DOCUMENT RESUME

ED 342 687 SE 052 988

AUTHOR Hambleton, Ronald K.; Bourque, Mary Lyn
TITLE The LEVELS of Mathematics Achievement: Initial Performance Standards for the 1990 NAEP Mathematics Assessment. Volume III: Technical Report.
INSTITUTION Aspen Systems Corp., Rockville, MD.
SPONS AGENCY National Assessment Governing Board, Washington, DC.
PUB DATE Nov 91
NOTE 446p.; For volumes 1 and 2, see SE 052 986-987.
AVAILABLE FROM NAGB Technical Report, 1100 L Street NW, Suite 7322, Washington, DC 20005-4013.
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC18 Plus Postage.
DESCRIPTORS *Achievement Rating; Construct Validity; Elementary Secondary Education; *Grade 4; *Grade 8; *Grade 12; *Mathematics Achievement; Mathematics Education; Mathematics Skills; Mathematics Tests; Measurement; National Programs; *Predictive Validity; State Programs
IDENTIFIERS *National Assessment of Educational Progress; Trial State Assessment (NAEP)

ABSTRACT

The National Assessment of Educational Progress (NAEP) is a congressionally mandated survey of educational achievement of American students in a variety of curriculum areas and of changes in that achievement over time. The National Assessment Governing Board (NAGB) has established new standards for reporting the results that determined three achievement levels: basic, proficient, and advanced. The basic level denotes partial mastery of the knowledge and skills fundamental for proficient work at each grade. Proficient, the central level, represents solid academic performance and demonstrated competence over challenging subject matter. The advanced level signifies superior performance beyond proficient. This book, volume III of the Initial Performance Standards for the 1990 NAEP Mathematics Assessment, describes the process of how these three levels were determined. The chapters include: Chapter 1: Executive Summary; Chapter 2: Overview to the Achievement Level-Setting Process; Chapter 3: Achievement Levels Methodology: Phase 1; Chapter 4: Analysis of Achievement Level Ratings: Phase 1; Chapter 5: Conclusions and Recommendations: Phase 1; Chapter 6: The Replication/Validation Study: Phase 2; Chapter 7: Analysis of Achievement Level Ratings--Validation/Replication; Chapter 8: Additional Topics; and Chapter 9: Conclusions and Recommendations. The bulk of the document is contained in the appendices that follow. These 14 appendices provide a detailed account of the development and validation of the established levels in a series of 85 related tables and documents. (MDH)

51

ED342687

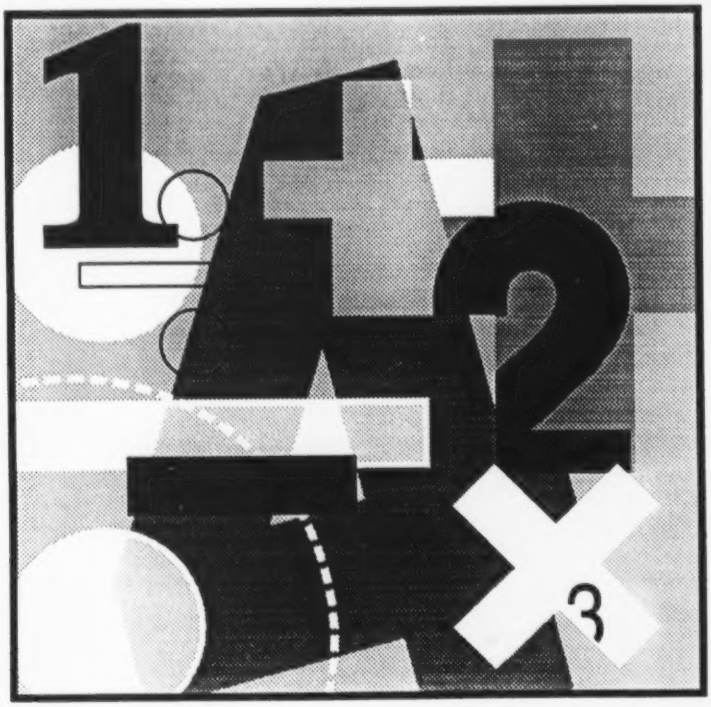
National Assessment Governing Board

The LEVELS of Mathematics Achievement

Initial Performance Standards for the 1990 NAEP Mathematics Assessment

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Volume III
Technical Report



SE053988

Prepared by Aspen Systems under contract with the National Assessment Governing Board

BEST COPY AVAILABLE

What is The Nation's Report Card?

THE NATION'S REPORT CARD, the National Assessment of Educational Progress (NAEP), is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, history/geography, and other fields. By making objective information on student performance available to policymakers at the national, state, and local levels, NAEP is an integral part of our nation's evaluation of the condition and progress of education. Only information related to academic achievement is collected under this program. NAEP guarantees the privacy of individual students and their families.

NAEP is a congressionally mandated project of the National Center for Education Statistics, the U.S. Department of Education. The Commissioner of Education Statistics is responsible, by law, for carrying out the NAEP project through competitive awards to qualified organizations. NAEP reports directly to the Commissioner, who is also responsible for providing continuing reviews, including validation studies and solicitation of public comment, on NAEP's conduct and usefulness.

In 1988, Congress created the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The board is responsible for selecting the subject areas to be assessed which may include adding to those specified by Congress; identifying appropriate achievement goals for each age and grade; developing assessment objectives; developing test specifications; designing the assessment methodology; developing guidelines and standards for data analysis and for reporting and disseminating results; developing standards and procedures for interstate, regional, and national comparisons; improving the form and use of the National Assessment; and ensuring that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias.

The National Assessment Governing Board

Richard A. Boyd, Chairman
Executive Director
Martha Holden Jennings Foundation
Cleveland, Ohio

Phyllis Williamson Aldrich
Curriculum Coordinator
Saratoga-Warren B.O.C.E.S.
Saratoga Springs, New York

David Battini
High School History Teacher
Cairo-Durham High School
Cairo, New York

Parris C. Battle
Education Specialist
Dade County Public Schools
Miami, Florida

Honorable Evan Bayh
Governor of Indiana
Indianapolis, Indiana

Mary R. Blanton
Attorney
Blanton & Blanton
Salisbury, North Carolina

Boyd W. Boehlje
Attorney
Gaass, Klyn, & Boehlje
Fella, Iowa

Linda R. Bryant
Dean of Students
Florence Reizenstein Middle School
Pittsburgh, Pennsylvania

Honorable Michael N. Castle
Commissioner of Education
Wilmington, Delaware

Honorable Naomi K. Cohen
Connecticut House of Representatives
Hartford, Connecticut

Chester E. Finn, Jr.
Professor of Education and Public Policy
Vanderbilt University
Washington, DC

Michael S. Glode
Wyoming State Board of Education
Saratoga, Wyoming

William Hume
Basic American, Inc.
San Francisco, California

Christine Johnson
Direcctor of K-12 Education
Littleton Public Schools
Littleton, Colorado

John Lindley
Principal
South Colby Elementary School
Port Orchard, Washington

Carl J. Moser
Director of Schools
The Lutheran Church - Missouri Synod
International Center
St. Louis, Missouri

John A. Murphy
Superintendent of Schools
Charlotte-Mecklenburg Schools
Charlotte, North Carolina

Mark Musick
President
Southern Regional Education Board
Atlanta, Georgia

Honorable Carolyn Pollan
Arkansas House of Representatives
Fort Smith, Arkansas

Honorable William T. Randall
Commissioner of Education
State Department of Education
Denver, Colorado

Thomas Topuzes
Senior Vice President
Valley Independent Bank
El Centro, California

Herbert J. Walberg
Professor of Education
University of Illinois
Chicago, Illinois

Diane S. Ravitch (Ex-Officio)
Assistant Secretary
and Counselor to the Secretary
U.S. Department of Education
Washington, DC

Roy Truby
Executive Director, NAGB
Washington, DC

BEST COPY AVAILABLE

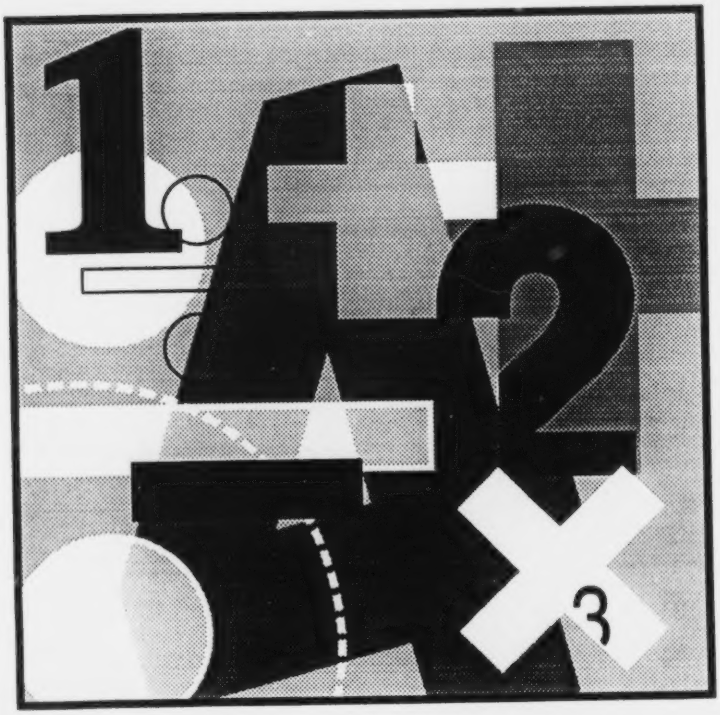
National Assessment Governing Board

The LEVELS of Mathematics Achievement

Initial Performance Standards for the 1990 NAEP Mathematics Assessment

Volume III
Technical Report

Ronald K. Hambleton
Mary Lyn Bourque



November 1991

Prepared by Aspen Systems under contract with the National Assessment Governing Board

National Assessment Governing Board
Richard A. Boyd
Chair

Mark D. Musick
Vice Chair

Michael S. Glode
Achievement Levels Committee Chair

FOR MORE INFORMATION:

Copies are available from the participating states, as well as from the National Assessment Governing Board, while supplies last. Write:

NAGB Technical Report
1100 L Street NW, Suite 7322
Washington, DC 20005-4013

or call (202) 357-6938.

Prepared by Aspen Systems under contract with the National Assessment Governing Board.

TABLE OF CONTENTS

CHAPTER PAGE

1. Executive Summary 1

    1.1 Summary 1
    1.2 National Assessment of Educational Progress 3
    1.3 The Governing Board 5
    1.4 The Policy Framework 6
    1.5 The Process of Setting Achievement Levels 8

2. Overview to the Achievement Level-Setting Process 17

    2.1 Introduction 17
    2.2 Vermont/Washington Study: Phase 1 17
    2.3 Validation/Replication Study: Phase 2 22

3. Achievement Levels Methodology: Phase 1 25

    3.1 Selection of Judges 25
    3.2 Technical Advisers and Reviewers 27
    3.3 Technical and Policy Evaluation 28
    3.4 Briefing and Training of Judges 29
    3.5 Item-Rating Tasks 30
    3.6 Content Descriptions of the Achievement Levels 32

4. Analysis of Achievement Level Ratings: Phase 1 35

    4.1 Introduction 35
    4.2 Overview of Results 35
    4.3 Attrition Prior to the Washington Meeting 40
    4.4 Explanation of the Adjustments in Tables 24, 25, and 26 42
    4.5 Achievement Levels for Content Categories and Abilities 43
    4.6 Item Appropriateness Ratings 44
    4.7 Correlations Between Expected and Actual Item Difficulty Values 44

5. Conclusions and Recommendations: Phase 1 47

    5.1 Adjustments to the Phase 1 Achievement Levels 47
    5.2 External Evaluations of the Level-Setting Process 48
    5.3 A Summary of the Problems 49
    5.4 Recommendations 51

TABLE OF CONTENTS (Continued)

CHAPTER PAGE

6. The Replication/Validation Study: Phase 2 53

6.1 Introduction 53
6.2 Replication/Validation Study 53
6.3 Selection of Judges 57
6.4 Training of Judges 58
6.5 Item Rating Tasks 58
6.6 Description of the Levels 59
6.7 Summary 59

7. Analysis of Achievement Level Ratings -Validation/Replication 81

7.1 Overview of Rounds One and Two Ratings 81
7.2 Comparisons of Achievement Levels Across Sites--Block Level 81
7.3 Final Round Achievement Levels 82
7.4 Evaluation of the Achievement Level-Setting Process 83
7.5 Evaluation of the Expanded Definitions 84
7.6 Additional Analyses 84
7.7 Comparison of Phase 1 and 2 Final Achievement Levels 86

8. Additional Topics 89

8.1 Introduction 89
8.2 Discrepancy in Time of Testing vs. End-of-Year Standards 89
8.3 Correction for Guessing 90
8.4 Estimating Variability 91

9. Conclusions and Recommendations 93

9.1 Summary 93
9.2 Advantages and Disadvantages of Phase 1 93
9.3 Advantages and Disadvantages of Phase 2 95
9.4 Recommendations for Future Efforts 96

10. References 101

7

TABLE OF CONTENTS (Continued)

CHAPTER PAGE

Appendix A: Panelists in Vermont/Washington, DC 103

Appendix B: Training Manual for Phase I 107

Appendix C: Briefing Materials and Meeting Agendas 123

Appendix D: Standard Setting Forms 131

Appendix E: Item Security Policy and Nondisclosure Form 171

Appendix F: Summary of Vermont and Washington Achievement Level Setting Data 177

Appendix G: Technical Memo 227

Appendix H: Panelists for Replication/Validation 241

Appendix I: Summary of Validation/Replication Achievement Level Setting Data 257

Appendix J: Setting Appropriate Achievement Levels for the National Assessment of Educational Progress Policy Framework and Technical Procedures 331

Appendix K: Replication/Validation Plan 375

Appendix L: Sample Trace Lines and Actual ICCs Used in Phase 1 385

Appendix M: Listing of Items in Grade-Level Pools in Order of p-Values 389

Appendix N: Acknowledgments 401

LIST OF TABLES

TABLE PAGE

1 Summary of Grade 4 Achievement Levels for the Total Item Pool 178

2 Summary of Grade 8 Achievement Levels for the Total Item Pool 178

3 Summary of Grade 12 Achievement Levels for Total Item Pool 179

4 Summary of Grade 4 Achievement Levels for the Reduced Item Pool 179

5 Summary of Grade 8 Achievement Levels for the Reduced Item Pool 180

6 Summary of Grade 12 Achievement Levels for the Reduced Item Pool ... 180

7 Summary of Grade 4 Third Round Achievement Levels, Reported for Groups (N=22) 181

8 Summary of Grade 8 Third Round Achievement Levels, Reported for Groups (N=22) 181

9 Summary of Grade 12 Third Round Achievement Levels, Reported for Groups (N=19) 182

10 Comparison of Estimated Average Difficulties at Round 3 for Items Which Were Common to Grades 4, 8 and 12 183

11 Comparison of Estimated Average Difficulties at Round 3 for Items Which Were Common to Grades 4 and 8 185

12 Comparison of Estimated Average Difficulties at Round 3 for Randomly Selected (50%) Common Items to Grades 8 and 12 186

13 Performance of the Average Student in the 1990 National Sample on Common Math Items 188

14 Summary of Average Item Performance and Achievement Levels on the Common Items After the Third Set of Ratings 188

15 Summary of Judges' Five Sets of Achievement Levels (Grade 4, 22 Judges) 189

16 Summary of Judges' Five Sets of Achievement Levels (Grade 8, 22 Judges) 190

17 Summary of Judges' Five Sets of Achievement Levels (Grade 12, 19 Judges) 191

LIST OF TABLES (Continued)

TABLE PAGE

18 Final 1990 NAEP Total Item Pool Mathematics Assessment Achievement Levels . . . 192

19 Descriptive Statistics on the Final Total Item Pool Mathematics Achievement Levels . . . 192

20 Summary of Confidence Levels of Judges in Setting Final Achievement Levels . . . 193

21 Return Rates of Judges to the Washington Meeting . . . 193

22 Comparison of the Demographic Composition of Judges at the Vermont and Washington Meetings . . . 194

23 Comparison of 3rd Set of (Vermont) Ratings for Judges "Not Present" and "Present" at the Washington Meeting . . . 194

24 Average (Adjusted) Grade 4 Item Achievement Levels . . . 196

25 Average (Adjusted) Grade 8 Item Achievement Levels . . . 199

26 Average (Adjusted) Grade 12 Item Achievement Levels . . . 203

27 Summary of Judges' Five Sets of Achievement Levels for the Reduced Item Pool (Grade 4, 22 Judges) . . . 207

28 Summary of Judges' Five Sets of Achievement Levels for the Reduced Item Pool (Grade 8, 22 Judges) . . . 208

29 Summary of Judges' Five Sets of Achievement Levels for the Reduced Item Pool (Grade 12, 19 Judges) . . . 209

30 Summary of Achievement Levels for Content Categories Based Upon (Adjusted) Fourth Round Ratings (Reduced Item Pool) . . . 210

31 Summary of Achievement Levels for Mathematics Abilities Based Upon (Adjusted) Fourth Round Ratings (Reduced Item Tool) . . . 211

32 Analysis of Grade 4 Item Appropriateness Ratings (N=10) . . . 212

34 Analysis of Grade 12 Item Appropriateness Ratings (N=8) . . . 219

35 Summary of Mean Item Appropriateness Ratings . . . 223

LIST OF TABLES (Continued)

TABLE PAGE

36 Correlations Between First, Second, and Third Round of Average Judges' Ratings of Expected Item p-Values and Actual p-Values 224

37 Summary of Grade 4 First Round Achievement Levels, Reported for Groups (N=22) 224

38 Summary of Grade 8 First Round Achievement Levels Reported for Groups (N=22) 225

39 Summary of Grade 12 First Round Achievement Levels Reported for Groups (N=19) 225

40 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 3, Judges = 30) 258

41 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 4, Judges = 25) 259

42 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 5, Judges = 30) 260

43 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 6, Judges = 26) 261

44 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 7, Judges = 22) 262

45 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 8, Judges = 33) 263

46 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 9, Judges = 29) 264

47 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 3, Judges = 27) 265

48 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 4, Judges = 31) 266

49 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 5, Judges = 31) 267

TABLE PAGE

50 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 6, Judges = 28) 268

51 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 7, Judges = 28) 269

52 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 8, Judges = 25) 270

53 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 9, Judges = 28) 271

54 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 3, Judges = 32) 272

55 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 4, Judges = 31) 273

56 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 5, Judges = 29) 274

57 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 6, Judges = 29) 275

58 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 7, Judges = 28) 276

59 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 8, Judges = 32) 277

60 Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 9, Judges = 29) 278

61 Summary of Grade 4 Achievement Levels at the Block Level for First and Second Ratings 279

62 Summary of Grade 8 Achievement Levels at the Block Level for First and Second Ratings 283

63 Summary of Grade 12 Achievement Levels at the Block Level for First and Second Ratings 287

64 Summary of Final Achievement Levels 291

LIST OF TABLES (Continued)

TABLE PAGE

65 Summary of Confidence Levels on the Final Ratings 292

66 Summary of Participant Evaluations of the NAGB Achievement Level Setting Process 294

67 Summary of Participant Evaluations of the NAGB Achievement Level Setting Process 297

68 Summary of the Achievement Level Review Results (Grade 4, N=66) 300

69 Summary of the Achievement Level Review Results (Grade 8, N=72) 301

70 Summary of the Achievement Level Review Results (Grade 12, N=73) 302

71 Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 4) 303

72 Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 8) 304

73 Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 12) 305

74 Analysis of Final Achievement Levels for Educators and Non-Educators 306

75 Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 4, 8, and 12 NAEP Test Booklets 307

76 Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 4 and 8 NAEP Test Booklets 308

77 Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 8 and 12 NAEP Test Booklets 309

78 Summary of Achievement Levels 311

79 Summary of Achievement Levels 312

80 Summary of Grade 4 Achievement Levels by Booklet and Round 313

LIST OF TABLES

TABLE | | PAGE
81 Summary of Grade 8 Achievement Levels by Booklet and Round 317

82 Summary of Grade 12 Achievement Levels by Booklet and Round 321

83 Grade 4 Achievement Levels by State 325

84 Grade 8 Achievement Levels by State 327

85 Grade 12 Achievement Levels by State 329

LIST OF FIGURES

FIGURE PAGE

1 Summary of Main Events in the Vermont/Washington Achievement Level-Setting Process: August and September 1990 18

2 Summary of Main Events in the Validation/Replication Achievement Level-Setting Process: March and April 1991 24

3 Mathematics Proficiency Corresponding to Each Achievement Level, By Grade: For 1990 NAEP Mathematics Assessment 61

LIST OF EXHIBITS

EXHIBIT PAGE

1 Levels of Mathematics Achievement for Grade 4 62

2 Levels of Mathematics Achievement for Grade 8 68

3 Levels of Mathematics Achievement for Grade 12 75

1. Executive Summary

1.1 Summary

For the past 20 years, the National Assessment of Educational Progress (NAEP), like virtually all nationally standardized tests in the United States, has reported results mainly in terms of average performance. Sometimes NAEP has announced the proportion of students who knew a certain fact or could demonstrate a certain skill. But it has shied away from saying clearly whether the average performance was good enough or whether there were any facts or competencies that students at any particular grade should be expected to know.

Under the legislation creating the National Assessment Governing Board (NAGB) in 1988, the Board was charged with responsibility for identifying "appropriate achievement goals for each...grade in each subject area to be tested under the National Assessment." The statute also gave the Board responsibility for "developing...standards for analysis plans and for reporting and disseminating [NAEP] results."

In May 1990, after wide public consultation and hearings, the Board unanimously adopted a policy to set achievement levels, defining what students should know and be able to do at different grades, on all NAEP assessments of fourth, eighth, and twelfth graders. Under the plan, three levels--BASIC, PROFICIENT, and ADVANCED--would be set for each grade and subject tested by NAEP. The Board voted that levels would first be set, as a trial, on the 1990 national assessment of mathematics. Using information from this experience, the Board resolved that starting in 1992, results for all future assessments would be reported initially and primarily in terms of achievement levels.

This Executive Summary describes the steps that Board committees and staff took, with the assistance of consultants, to prepare recommended achievement levels for Board action. One set of proposed levels--with achievement levels and descriptions of each level--was presented to the Board in November 1990. These achievement levels were based on meetings held in August, September, and November of 1990 in Vermont and Washington involving 63 educators and noneducators from across the country.

After two public hearings, reports by outside evaluators, and input from several other organizations and individuals, the Board directed staff to conduct a validation/replication study. This second study involved 211 persons--of whom more than 80 percent were classroom teachers--who took part in meetings in March and April of 1991 in California, Connecticut, Florida, and Michigan in cooperation with state education departments. Officials of these departments helped to develop plans for the meetings and to invite participants.

A second set of achievement levels was prepared based on the validation/replication study, including descriptions and illustrative items for each achievement level. The text of the proposed descriptions was written by a panel of mathematics experts headed by John A. Dossey of Illinois State University. The expert panel also prepared revised descriptions for achievement levels based on the work of the participants.

Generally, the achievement levels recommended by the validation/replication panels were somewhat lower than those proposed by the Vermont/Washington advisory group. There was virtually no change for twelfth grade ADVANCED. At eighth grade BASIC the drop was most substantial. However, this change placed the percentage of expected correct answers at grade 8 BASIC more in line with the percentage expected for the BASIC levels at grades eight and twelve. This change also created a clear distinction between the BASIC levels at the eighth and twelfth grade which had not been present in the achievement levels set by the Vermont/Washington panel.

On May 10, 1991, after reviewing the options before it and hearing reports from the lead technical consultant and the evaluators, the Board adopted achievement levels based on the validation/replication study for reporting and interpreting results of the 1990 mathematics assessment.

1.2 National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) is currently the only nationally representative and continuing assessment of what American students know and can do in various academic subjects. Mandated by Congress, the assessments have been conducted on a nationwide sample survey basis since 1969. Subjects tested have included reading, mathematics, science, writing, U.S. history, and geography. At various times, assessments have also been conducted of civics, computer competence, art, music, literature, and health.

In 1990, as authorized by Congress, NAEP collected comparable state-by-state data for the first time on a voluntary trial basis in eighth grade mathematics. Thirty-seven states, the District of Columbia, and two territories participated in this program which involved testing a representative cross section of about 2,500 students per state.

In 1992, trial state assessments will be conducted in fourth and eighth grade math, and in fourth grade reading. Nationwide sample testing has been authorized every 2 years in three to five subjects in grades 4, 8, and 12, and at ages 9, 13, and 17. About 220,000 students participated in the 1990 assessments, including about 80,000 eighth graders who were part of the state level samples. By law, NAEP cannot report data below the state level, i.e., on individuals, schools, or school districts.

At present, the assessment is conducted by the Educational Testing Service (ETS) under contract to the National Center for Education Statistics of the U.S. Education Department. The Commissioner of Education Statistics is responsible for administering the program under policy guidance of the National Assessment Governing Board.

Since 1983, in an effort to improve public understanding of NAEP results, ETS has described the types of skills that can be performed by students using a set of arbitrarily chosen points on the NAEP score-reporting scales. These points have been based on the distribution of test results, not on any judgment about what students ought to know or be able to do. Under this system, NAEP data for each subject are reported on a common, empirically derived cross-grade scale that spans grades 4, 8, and 12. Each scale has a mean score of 250. Each 50-point interval represents (approximately) one standard deviation--a measure of variation in test scores--across all students in all three grades tested. The cluster of skills that differentiates each major level is determined by looking at the patterns of right and wrong answers after the assessment is administered. Based on test questions that differentiate students at the 150, 200, 250, 300, and 350 levels, descriptions are written characterizing the knowledge and skills which students at each of these five anchor points are most likely to have.

Although the ETS proficiency levels have been helpful in explaining NAEP results, they are based solely on statistical distributions of test performance. Thus, they provide only limited guidance for determining whether students have mastered challenging subject matter or have acquired the knowledge and skills needed to advance in school or move on successfully to college and adulthood.

The National Assessment Governing Board believes that defining what performance on NAEP ought to be through a careful, broadly based judgmental process will greatly enhance the assessment's central function as a yardstick of educational achievement by American students.

1.3 The Governing Board

The National Assessment Governing Board (NAGB) was created in 1988 under Public Law 100-297 to set policy for the National Assessment of Educational Progress. The 24-member Board is composed of a broadly representative group of state, local, and federal officials; educators; and members of the public. It is appointed by the Secretary of Education in categories prescribed by law from among nominees proposed by the Board itself.

In addition to identifying appropriate achievement goals for each grade and subject tested, the Board develops assessment objectives and test specifications; designs the assessment methodology and standards for reporting results; selects subject areas to be assessed, in addition to those specified by law; and has final authority on the appropriateness of test items. The Board also has general responsibility "to improve the form and use of the National Assessment."

According to the statute, "In the exercise of its functions, powers, and duties, the Board shall...be independent of the Secretary and the other offices and officers of the Department of Education." The legislation creating the Governing Board was based in part on recommendations made in 1987 by a study group on NAEP, chaired by Lamar Alexander, then governor of Tennessee, who became Secretary of Education in March 1991. The vice chairman and study director was H. Thomas James, president emeritus of the Spencer Foundation.

The Alexander-James study group stated in its report that: "The governance and policy direction of the national assessment should be furnished by a broadly representative [Board] that provides wisdom, stability, and continuity; that is charged with meshing the assessment needs of states and localities with that of the nation; that is accountable to the public--and to the federal government--for stewardship of this important activity; but that is itself buffered from manipulation by any individual, level of government, or special interest within the field of education."

As prescribed by law, the Board should include two governors or former governors of different political parties, two state legislators of different parties, two chief state school officers, one local school superintendent, three classroom teachers, one state and one local school board member, two testing and measurement experts, two school principals, two curriculum specialists, one business representative, one representative of private schools, three general public members, and the Assistant Secretary for Educational Research and Improvement (ex-officio).

1.4 The Policy Framework

Although the Board was authorized to identify "appropriate achievement goals" on NAEP long before national education goals were formulated, NAGB kept the national goals in mind when framing its policy. In particular, the Board considered the need to make NAEP more useful in tracking progress toward Goal Three, which states that "By the year 2000, American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter, including English, mathematics, science, history, and geography." The phrase "having demonstrated competency in challenging subject matter" was incorporated as the main defining language of the Board's general description of the proficient level for each grade. Six national goals were set by the President and the nation's governors in September of 1990.

According to the Board resolution of May 11, 1990, NAGB intended to establish three achievement levels for each grade and subject tested under NAEP. It will report the proportion of students who meet or exceed each achievement level. The levels will have clear distinctions among them, will be illustrated by representative sample items, and will be coherent and consistent over grades 4, 8, and 12 in the NAEP assessment.

The generic definitions of the achievement levels prepared by NAGB are as follows:

(a) BASIC. This level denotes partial mastery of knowledge and skills that are fundamental for proficient work at grades 4, 8, and 12. For twelfth grade, this will be higher than minimum competency skills (which normally are taught in elementary and junior high schools) and will cover significant elements of standard high school-level work.

(b) PROFICIENT. This central level represents solid academic performance for grades 4, 8, and 12. It will reflect a consensus that students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. For twelfth grade, the proficient level will encompass a body of subject-matter knowledge and analytical skills, of cultural literacy and insight that all high school graduates should have for democratic citizenship, responsible adulthood, and productive work.

(c) ADVANCED. This higher level signifies superior performance beyond proficient grade-level mastery at grades 4, 8, and 12. For twelfth grade, the advanced level will show readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, it may be based in part on international comparisons of academic achievement or it may be related to advanced placement and other college placement exams.

NAGB applied these definitions in setting achievement levels on the 1990 national assessment of mathematics. The current plan is to define achievement levels on the new NAEP tests of reading and writing for 1992, and in science, U.S. history, and geography for 1994. It will also reset the mathematics achievement levels in 1992, since the 1990 work on the mathematics achievement levels was only a trial.

1.5 The Process of Setting Achievement Levels

Since this achievement level-setting effort was perhaps the largest and most important ever in American education, NAGB felt it must be open to public scrutiny and input and that every effort should be made to secure technical consultation.

Appointment of Advisory Panel--June 1990. NAGB appointed a panel of 63 judges. About 70 percent of the panel members were educators, representing subject-area teachers, college mathematics instructors, principals, and state and district curriculum specialists; 30 percent were noneducators representing employers, civic group representatives, and interested citizens; and 20 percent were minority group members. Gender and geographical representation was also considered when making appointments. Panelists came from schools from New York to California, from the inner-city schools of Detroit and Chicago, and from the suburbs of Winnetka, Illinois, and Huntington, Connecticut. They represented every part of the country and nearly every subgroup of the nation's population.

Vermont Meeting--August 16-17, 1990. Achievement level setting is a judgmental process. The meeting in Essex Junction, Vermont, provided background and a framework for the panel members to share their judgments. The meeting proceeded as follows:

1. Judges received training about the process.
2. Panelists met in four small, heterogeneous groups at each grade level--4, 8, and 12. The groups were given the item pool from the 1990 math assessment. Each judge was asked to make a first round of ratings, indicating what proportion of students at each achievement level should answer each particular question correctly. These ratings were aggregated over items first to determine achievement levels for each

judge and then later averaged over judges to produce a recommended percentage correct score for each achievement level at each grade.

3. The groups were given information on how students actually performed on each question during the 1990 testing. Each judge then did a second round of ratings, with little or no group discussion. Having performance information caused very little overall change in the ratings.
4. The judges completed a third round of ratings. This time the judges discussed with others in their group their first two rounds of ratings. Then they provided their third set of item ratings.
5. The results of the third round of ratings were shared with the judges from all three grades. Unfortunately, the additional two steps in the process--designed to achieve consistency and coherence in the achievement levels--could not be completed because time was not available. These last two steps involved discussions among all judges at a particular grade level and discussions among judges across grade levels.

Post-Vermont Meeting

6. Revisions were made in some procedures based upon discussions with the technical advisory committee on achievement-level setting. Two concerns were given special attention:
    (1) Making sure judges had a clear understanding of the Board's general definitions of BASIC, PROFICIENT, and ADVANCED.
    (2) Ensuring that judges based their ratings on the difficulty of test items and their importance in showing mastery rather than on whether an item or item format was appropriate for inclusion in NAEP.

7. Analyses of the first three rounds of ratings were prepared.

First Washington Meeting--September 29-30, 1990

8. Thirty-eight of the 63 judges reconvened in Washington.
9. The judges discussed the definitions of BASIC, PROFICIENT, and ADVANCED in order to clarify them.
10. Judges completed a fourth round of rating individual questions.
11. The judges met with others at their own grade level, and later in groups that included panelists from all three grades, to discuss the consistency and coherence of the recommended levels.
12. Judges made a fifth round of ratings giving the overall percentage correct that should be required to reach each achievement level for their grade.
13. Judges completed an evaluation form expressing their confidence levels in their own final ratings.

Second Washington Meeting--November 12-13, 1990

14. Eleven judges wrote descriptions of the three achievement levels for each grade based on analyses of individual item ratings and average expected percent correct (adjusted round four ratings) derived from judges' rating forms at earlier meetings. Sample items were selected to illustrate each achievement level.
15. The text of the final recommendations was sent to all panel members for approval. In written replies, 45 expressed approval; 8 disagreed in whole or in part; and 10 did not respond.

NAGB Board Meeting in Atlanta--November 16-17, 1990

16. The recommended achievement levels were presented to NAGB at this meeting. The Board also heard comments from the project's lead technical consultant, Ronald K. Hambleton of the University of Massachusetts at Amherst, and from the lead evaluator, Daniel Stufflebeam of Western Michigan University. The evaluators recommended moving forward to completion but cautioned NAGB to proceed slowly enough to allow extensive public input.

Public Comment--November 1990 to January 1991. Oral and written testimony was received from about 30 persons and organizations at public hearings in Washington on November 26, 1990 and January 8, 1991. Comments were about evenly divided. Proposed achievement levels were praised as embodying strong, useful standards for mathematics achievement. They were also criticized as having been developed too quickly and on an item pool not specifically designed to accommodate the achievement levels. Constructive but negative evaluations of the process and results were also received from the panel of independent evaluators and from the NAEP Technical Review Panel.

NAGB Board Meeting in Washington--March 1-2, 1991. In response to concerns from several sources, the Board adopted a validation/replication plan outlining procedures to obtain advice from panels of experts across the country--primarily teachers--on what the achievement levels should be.

Validation/Replication Process--March and April 1991. All-day meetings were conducted in four states in different regions of the country to receive recommendations from the participants composed mostly of mathematics teachers. The process was intended to gather a broad cross-section of informed opinion in a carefully organized way. Participants were asked to give their opinion based on their personal experience and viewpoint of what students at different levels of achievement should be able to do. The results of the four meetings were aggregated to produce recommendations for the Board--expressed as the percentage of questions that students should answer correctly to reach the BASIC, PROFICIENT, and ADVANCED levels for each grade.

To ensure uniformity among the meetings, the same format was followed at all the meetings. Mary Lyn Bourque, NAGB Assistant Director for Psychometrics, conducted the meetings, which were held in Cromwell, Connecticut; Lansing, Michigan; Los Angeles, California; and Tampa, Florida. Because of low attendance in Lansing, a second session for Michigan was held later in Detroit. The state departments of education assisted in arranging the meetings and in assembling participants according to criteria established by NAGB. At each site, there was a cross section of teachers from urban, suburban, and rural schools with a range of years of experience who had worked with children of varying ability levels. Almost all of the participants came from the four participating states although a few were from nearby states:

17. Of 211 participants, 77 percent were white, 15 percent black, 4 percent Hispanic, and 2 percent Asian. Sixty percent were female and 40 percent male. Forty-three percent said they taught or worked in an urban or mostly urban community, 42 percent in a suburban community, and 15 percent in a rural or mostly rural community.
18. Of the 25 noneducators in the validation/replication groups, about half were representatives of business and industry and half were school board members and parent representatives.

19. Of the teachers, 49 percent said they taught mostly average mainstream students; 27 percent mostly above-average students; 19 percent mostly below-average students; and 5 percent mostly students with special needs.

The format of the meetings was as follows:

1. After an introductory briefing, partly through videotape prepared by Ronald Hambleton, each judge was given a NAEP test booklet. The booklets were distributed according to the standard matrix sampling NAEP design. Thus, each rater had three-sevenths of the test blocks (45 to 65 questions) for the grade he or she was considering.
2. After the judges had worked through each problem and had checked the answer, they made their first ratings. For each test item, they were asked to apply the definitions approved by the Board and write down what proportion of students who had just reached the BASIC, PROFICIENT, and ADVANCED levels should answer each question correctly.
3. Judges were then given item-by-item results of the 1990 NAEP mathematics assessment, showing the proportion of students that actually answered each question correctly. They were asked to make a second rating which allowed them to modify their initial judgment if they wished, in light of the test results they had received.
4. Overall, these second round ratings tended to be slightly lower than the first ratings by an average of about 3 percentage points.
5. Staff averaged the expected percentage correct for each question and calculated the overall percentage correct for each achievement level that had been recommended in both rounds one and two. This information was shared with judges.

6. Judges discussed these averages in small groups composed of raters in their own grades and in other grades, allowing them to consider the issues of coherence (across grades) and consistency (within grades) of the proposed achievement levels.

7. Each judge made a final rating in terms of the overall percentage correct for each achievement level at the grade being considered. These were averaged to produce the final recommendations of the validation/replication panels in terms of the expected percentage correct for BASIC, PROFICIENT, and ADVANCED achievement at each of the three grades. Nearly all of these figures were slightly higher than the recommendations calculated after round two, with an average increase of about 2 percentage points.

8. Even though there were differences of opinion among the judges, the relatively slight variations in the round-to-round averages indicated a high degree of consistency in their ratings.

9. To prepare the written descriptions of achievement levels, an analysis was made of the judges' item-by-item ratings in round two. This identified questions that the judges felt distinguished between the achievement levels. The panel of mathematics experts headed by John Dossey used this information to prepare the written descriptions and to select sample items (from among those available for public release) to illustrate each proposed level.

NAGB Board Meeting in Washington--May 10-11, 1991. The achievement level descriptions and recommended percentage correct for each level, as prepared through the validation/replication process, were approved by the Governing Board on a 19 to 1 vote. After separate reports describing and evaluating the total process were provided by the head technical consultant (Ronald K. Hambleton) and the three-person evaluation team (Daniel Stafflebeam, Richard Jaeger, and Michael Scriven), the resolution indicated that minor changes could be made as a result of editing and further analysis. No changes in the achievement levels were made. Although the levels were used in reporting and interpreting results of the 1990 NAEP mathematics assessment, they will be subject to review before being applied to the 1992 results.

2. Overview to the Achievement Level-Setting Process

2.1 Introduction

The technical portion of NAGB's efforts to set achievement levels began in May of 1990 when the authors were invited to prepare a handbook for judges describing the proposed achievement-level setting process. At that time, the first author also agreed to coordinate the Essex Junction, Vermont, meeting where the achievement levels would be set. What began as a four-day and later extended to an eight-day contract became an intensive one-year study to design the achievement-level setting process, to collect and analyze the item ratings data, to participate in various planning and review sessions, and then to respond to reactions to the process itself. In this section of the report, an overview to the first and second studies to set achievement levels, the Vermont/Washington initiative and the validation/replication initiative, will be described. Chapters 3 and 4 describe the details of the process and the results for the Vermont/Washington initiative. Chapters 6 and 7 provide the corresponding information for the Validation/Replication initiative.

2.2 Vermont/Washington Study: Phase 1

Figure 1 contains the 28 steps carried out during the Vermont/Washington phase of the project. Basically, the plan required the judges to make "Angoff-like" (1971) ratings for the marginally BASIC, PROFICIENT, and ADVANCED student at the grade level to which they were assigned (grades 4, 8, or 12). The judges were asked to specify the probability with which the minimally capable student at each of the three levels should answer each question in the 1990 NAEP mathematics assessment. Judges provided five sets (or five rounds) of ratings as follows:

Figure 1-- **Summary of Main Events in the Vermont/Washington Achievement Level-Setting Process: August and September 1990**

Pre-Vermont Meeting

1. Selected 63 judges and provided them with background materials such as 1990 mathematics objectives, sample test items and the NAGB report on achievement levels.

Vermont Meeting (August 16-17, 1990)

2. Convened 63 judges, NAGB staff, evaluators, and numerous observers in Essex Junction, Vermont.
3. Provided an overview of the goals of the achievement level setting process.
4. Provided technical training in the modified Angoff method.
5. Completed the first round of ratings. Judges at each grade level were organized into heterogeneous groups of 5 and 6. Definitions of marginally basic, proficient, and advanced students were discussed first, and then judges provided their item ratings. Discussions among the judges did not take place during the item rating process.
6. Completed the second round of ratings. Judges were given normative data (p-values and trace lines). After these data were explained, judges completed the second round of item ratings. Again, little or no discussion took place among the judges.
7. Completed the third round of ratings. Within each of the groups, judges participated in a discussion of their first and second round of item ratings. Low and high ratings for each item were discussed, along with other points about the item (e.g., shortcomings of the item, plausibility of distractors, format), and then a third round of item ratings was provided. Typically, discussion on an item took place, then judges provided a third rating, and then discussion moved to the next item.
8. Shared the results of the third round of ratings with the total group of judges.

Pre-Washington Meeting

9. Revised some of the procedures based upon informal discussions with NAGB staff, the formative evaluation team, and the technical advisory committee on standard setting. Three concerns were given special attention:
   - Clarifying the definitions;
   - Judging item appropriateness; and
   - Insuring separation of item difficulty and item appropriateness in the item ratings.
10. Conducted various analyses of the item ratings and prepared tables (see, for example, Tables 1 to 17, minus the round four and five results).

Figure 1-- **Summary of Main Events in the Vermont/Washington Achievement Level-Setting Process: August and September 1990 -- Continued**

Washington Meeting (September 29-30, 1990)

11. Reconvened 38 of the 63 judges in Washington.
12. Conducted a two-hour discussion of the definitions of basic, proficient, and advanced students.
13. Completed the fourth round of ratings. Here, judges were instructed to focus on the difficulty of items for the marginally basic, proficient, and advanced students. Item appropriateness was not to be considered in these ratings.
14. Completed an item appropriateness rating form.
15. Presented a complete set of analyses of item ratings for rounds one to three, and summary results for round four. Inconsistencies in the ratings, some of which were identified with the common items, were highlighted (see tables 10 to 14).
16. Conducted separate meetings of grade 4, 8, and 12 judges to consider the results, with an emphasis on consistency and coherence of the achievement levels.
17. Conducted two parallel meetings of grade 4, 8, and 12 judges (50% in each meeting) to consider the results, with an emphasis on consistency and coherence of the achievement levels.
18. Conducted a meeting of the total group of judges to consider the results with an emphasis on consistency and coherence of the achievement levels.
19. Collected a fifth and final set of ratings. Judges completed a one-page rating form in which they provided their final ratings and their confidence levels in these ratings.
20. Reported to the total group of judges the recommended achievement levels based upon the fifth and final ratings of the 38 judges (see table 18).

Post-Washington Meeting

21. Participated in a meeting with NAGB and ETS staff and the technical advisory committee on achievement level setting, and four actions were recommended:
    - Adjust round five data to reflect the views, to the extent possible, of judges who were unable to be in Washington on September 29 and 30.
    - Revise the achievement levels by removing higher order thinking skills and estimation items.
    - Substitute medians for means in arriving at the achievement levels.
    - "Smooth" the achievement levels to achieve more consistency and coherence.
22. Proposed preliminary achievement levels (see step 20) to NAGB (uninfluenced by step 21).
23. Transformed the achievement levels from step 20 to NAEP reporting scale and preliminarily determined their coherence.
24. Presented the achievement levels from step 20 to NAGB.
25. Responded to some of the reporting and analysis suggestions from reviewers and prepared the December 7, 1990, report of statistics (included 32 tables).

Figure 1-- **Summary of Main Events in the Vermont/Washington Achievement Level-Setting Process: August and September 1990 -- Continued**

Final Steps

26. Sought technical advice from the groups who participated at step 21 on a proposed set of minor revisions to the achievement levels. (For the results see the memo in appendix G)
27. Made revisions to the achievement levels and ETS mapped the levels onto the NAEP reporting scale using item response theory (IRT) methods and equations. Reviewed the achievement levels for coherence.
28. Made necessary revisions and presented final recommended achievement levels to NAGB.

---

1. Judges worked through the items independently and provided item ratings. They had access to the scoring key and knew (or could find out if they wanted) the objectives the items measured.
2. Judges were provided with each item difficulty level (p-value) for the 1990 sample of students and an "item-block score regression line" (something crudely approximating an "item characteristic curve" in which test scores at the block level served as the independent variable) which reflected the increase in actual item performance for students with different math abilities. (See appendix L for an example.)
3. At each grade level, four heterogeneous groups (to the extent possible) of five or six judges were formed to review independent ratings at steps 1 and 2, to discuss their differences, and then to provide a third set of ratings. The four groups were kept independent of one another and, therefore, served as four replications of the process at each grade level.
4. The total group of judges worked to further clarify the definitions of BASIC, PROFICIENT, and ADVANCED students, and then, after being reminded to base their ratings solely on their perceptions of item difficulty (independent of item appropriateness), they provided another (fourth) set of item ratings.
5. Judges were provided with a complete analysis of the first three sets of item ratings and the summary results (i.e., achievement levels) from the fourth set of item ratings. Then, all of the judges at each grade level met to discuss the complete set of results up to that point. Next, two "parallel" groups of judges across the grade levels met to discuss the results, and then all of the judges met to discuss the results. Finally, judges provided their fifth and final set of achievement levels on a scale of zero to

100 percent. They also provided ratings of their confidence levels in the achievement levels they had set.

The steps described above are what actually happened. Steps 1 through 3 went as originally planned for the Essex Junction, Vermont meeting, although they took more time to complete than had been planned. Unfortunately, there was insufficient time to complete steps 4 and 5 in the original plan. NAGB decided to reconvene the judges in Washington in late September of 1990 to complete the process. Since extra time was available at the Washington meeting, step 4 was revised from the original plan to respond to a number of methodological problems (i.e., confusion over definitions and the item ratings process itself) that had arisen in steps 1 to 3. The following factors contributed to the time problem at the Vermont meeting:

- Many judges wanted answers to questions that were not directly related to the achievement level-setting process.
- Many judges wanted to address their own issues and concerns prior to initiating the process.
- About two weeks prior to the meeting at the request of ETS, the item pools were expanded to include the estimation and higher order thinking skills items. This resulted in additional time required to complete the item rating task.

2.3 Validation/Replication Study: Phase 2

For reasons that will be described in chapter 5, NAGB made the decision to go ahead with a second study, referred to here as the validation/replication study. The goals of the study were to:

- Collect additional achievement level-setting data to validate the earlier results, or to improve upon them, if possible.

- Improve (without totally redesigning) the achievement level-setting process, by responding to some of the flaws noted in the earlier work.

Figure 2 describes the 10 steps in the validation/replication study. There were a number of differences in the methodology of this second study including the following:

- Reduced the time from four days to one day.
- Reduced the amount of advanced background materials to participants.
- Focused on classroom teachers (more than 80 percent).
- Reduced the item rating task (from 150-200 items to about 50 items per image; and from five rounds to three rounds).
- Simplified the item statistics information (no trace lines were used).
- Reduced the time for grade and across grade discussions. (This was necessary because few participants rated the same items and because there was no time to provide extensive feedback on item rating results).
- Standardized the (main) training by using a 35-minute videotape.
- Substantially increased the number of participants, from 39 to 211, though the amount of item ratings data collected from each judge was substantially reduced to three-sevenths of the reduced item pool (reduced by deleting EST and HOTS items).

All in all, the four meetings were conducted smoothly, and the majority of participants felt very positive about the experience and the results.

Figure 2-- Summary of Main Events in the Validation/Replication Achievement Level-Setting Process: March and April 1991

1. Proposed basic one-day design, received feedback from numerous groups and individuals, and revised plans.
2. Selected four sites and 50 to 60 participants per site.
3. Prepared a 35-minute video describing the achievement level-setting process, which was used during the training of participants.
4. Conducted a field test of the one-day meeting in the District of Columbia area and made minor revisions as necessary.
5. Distributed advanced materials to participants.
6. Conducted the one-day meetings which included:
    - An overview of the process;
    - Independent item ratings;
    - Independent item ratings with item statistics; and
    - Discussions with participants (who rated the same booklets) and then brief grade and across grade discussions.
7. Analyzed the main results and prepared tables.
8. Presented the results to NAGB on May 10, 1991.
9. Conducted additional analyses of the results (e.g., open-ended survey results) and updated results; extended tables.

3. Achievement Levels Methodology: Phase 1

3.1 Selection of Judges

The selection of judges for the achievement level-setting meeting in Vermont was initially implemented by contacting the major national organizations listed below and requesting that they nominate members of their organization to serve on the panels. The following organizations were initially contacted for nominees and alternates:

American Federation of Teachers
Association of School Assessment Programs
Association of School Supervisors of Mathematics
Association for Supervision and Curriculum Development
College Entrance Examination Board
Council for American Private Education
Council for Basic Education
Council of Chief State School Officers
Educational Testing Service
National Academy of Sciences, Mathematical Sciences Education Board
National Alliance of Business
National Association of Elementary School Principals
National Association of Secondary School Principals
National Association of State Boards of Education
National Association of Test Directors
National Catholic Education Association
National Council of Teachers of Mathematics
National Education Association
National School Boards Association
National Parent Teachers Association
United States Armed Forces

Nominees had to meet the criteria established by the Board in its policy paper. More than 20 of the organizations responded by recommending about 300 individuals for consideration by the Board.

As a matter of policy, the Board wanted individuals with expertise in the education of students in grades 4, 8, and 12; specifically, experience in the assessment of students' achievement in the area of mathematics and general knowledge of the typical mathematics achievement of students of the ages and grades under consideration. There should be overlapping membership between the achievement level-setting panel members and the original consensus groups convened in 1988 to articulate the 1990 mathematics assessment framework. Likewise, there should be special consideration given to nominees from states who were participating in the 1990 Trial State Assessment. The panel should have gender and racial/ethnic representativeness, and about one-third of the members should represent noneducators.

About 70 individuals were invited and agreed to participate in the meeting held in Essex Junction, Vermont, on August 15-16, 1990. Sixty-three persons representing 29 states and the District of Columbia attended the Essex Junction, Vermont, meeting and participated in the level-setting process. States represented in the meeting included:

Arizona
Arkansas
California
Connecticut
District of Columbia
Florida
Georgia
Illinois
Iowa
Kansas*
Maryland
Massachusetts*
Michigan
Minnesota
New Hampshire
New York
North Carolina
Ohio
Oklahoma
Oregon*
Pennsylvania
South Carolina*
Tennessee*
Texas
Utah*
Vermont*
Virginia
Washington*
Wisconsin
Wyoming

* States not participating in the Trial State Assessment Program

The panel was composed of 30 (48 percent) males and 33 (52 percent) females. The racial/ethnic composition was 83 percent majority and 17 percent minority, which included 8 blacks, 1 Asian, 1 Hispanic, and 1 Native American. About 30 percent of the panel were noneducators representing business and industry, the military, government service, parents, and the general public. Each panel member was assigned to a particular grade level for reviewing the item pool based on their stated preference or background. This resulted in 22 judges at grades 4 and 8 and 19 at grade 12.

Because insufficient time had been allocated for completing all the tasks at the Vermont meeting, a second meeting was held six weeks later in Washington, DC. Because the only available dates were exactly prior to the close of the 1989 fiscal year, which coincided with the observance of religious holidays, only 39 of the 63 members participated in the second meeting. This resulted in having only 11 judges at grade 4, 9 at grade 12, and all 19 of the original 22 judges at grade 8 in attendance.

3.2 Technical Advisers and Reviewers

Throughout the process for setting achievement levels the Board and its staff sought to obtain the best possible technical advice available from a variety of individuals. A Technical Advisory Committee on Standard Setting (TACSS) was formed that met whenever important methodological issues arose. Serving on the TACSS during part or all of the committee's deliberations were Richard Jaeger from the University of North Carolina at Greensboro; Robert Forsyth from the University of Iowa; Edward Haertel from Stanford University; Ronald K. Hambleton from the University of Massachusetts, who also served as the principal consultant for the project; and Eugene Johnson and Ina V.S. Mullis, both from the Educational Testing Service, the current NAEP operations contractor.

During its deliberations, the TACSS advised on such issues as: (1) mapping the achievement levels onto the NAEP scale; (2) interpretation and display of item data using the achievement levels; (3) appropriate data analyses to be conducted after the Vermont meeting; (4) using the judges' data to describe the knowledge and skills needed by students at each achievement level; (5) suggestions for identifying appropriate sample items for each level; and (6) other measurement concerns raised by stakeholder groups throughout the process.

In addition to the TACSS, several professionals in the measurement and mathematics fields reviewed training materials to be used in Vermont to ensure their technical accuracy and general clarity. Reviewers included Ronald Berk, Johns Hopkins University; John Carroll, Chapel Hill, North Carolina; Walter Denham, California Assessment Program; Jeremy Finn, SUNY Buffalo; Edward Haertel and Ingram Olkin, Stanford University; Sylvia Johnson, Howard University; Ina Mullis, Educational Testing Service; Eugene Owen and Gary Phillips, National Center for Education Statistics; and John Tukey, Princeton, New Jersey.

3.3 Technical and Policy Evaluation

Because the policy and technical framework document called for a formal evaluation of the process for setting achievement levels, the Board engaged the services of the Evaluation Center at Western Michigan University. The evaluation team included Richard M. Jaeger, professor and director of the Center for Educational Research and Evaluation of the University of North Carolina at Greensboro; Michael Scriven, consulting professor at Stanford University and adjunct professor at Western Michigan University; and Daniel L. Stufflebeam professor and director of the Evaluation Center at Western Michigan University. Sally Veeder served as administrative assistant and project secretary.

While the evaluation team worked collaboratively and produced a jointly signed report, each member also provided leadership for the team regarding a particular feature of the standard setting process. According to the evaluation proposal, Richard Jaeger examined particularly the modified Angoff methodology and its application in this specific setting; Michael Scriven examined policies and definitions, which formed the basis for the policy framework of the project; and Daniel Stufflebeam identified relevant concerns of stakeholders and examined the overall standard setting project.

The anticipated completion date for the evaluation was November, but because the work of the Board was still continuing at that time due to unforeseen circumstances, the evaluation team presented an interim report to the Board at its November 15-16 meeting in Atlanta. This was phase I of the work. The evaluation team continued its work through the spring of 1991 and presented a second interim report to the Board in May at its meeting in Washington, DC, based on phase II of the work. A draft final evaluation report was submitted on August 13, 1991 and the final evaluation report, phase III, was submitted on August 26, which contained the final recommendations of the evaluation team.

3.4 Briefing and Training of Judges

Since the judges were not equally familiar with the National Assessment program and the achievement level-setting initiative of the Board, and since the group was fairly heterogeneous in its areas of expertise, a variety of background reading materials was provided to the judges prior to their sitting on the panels. Briefing materials included the 1990 NAEP objectives; the NCTM curriculum standards; a training handbook for judges; and sample item-sets from the College Entrance Examination Board, the International Baccalaureate program, the American College Testing program, and the Advanced Placement program. These sample item-sets were meant to demonstrate what the Board had in mind when it proposed an ADVANCED level for one of the standards. The item-sets also reflected the expectations of major testing programs in which American students compete on a regular basis.

The training handbook, contained in Appendix B and developed by Ronald Hambleton, described the background and rationale for the judges' work. It also provided a detailed description of the achievement level-setting method; working descriptions of BASIC, PROFICIENT, and ADVANCED students; a practice achievement level-setting exercise; and step-by-step instructions for the judges.

The handbook was prepared to reflect the Board's policy on achievement levels. In the training in Vermont, as in the handbook itself, there was no attempt to elaborate the generic definitions for BASIC, PROFICIENT, and ADVANCED. The training materials were designed to provide the judges with insight into the Board's thinking, so that they could make appropriate judgments about the item pool and arrive at the achievement levels--levels which reflected the very best professional judgment of math educators, noneducators, and the general public.

3.5 Item-Rating Tasks

A major modification of the Angoff method for this project was in the item-rating tasks required of the judges. Typically, in other Angoff procedures documented in the literature (Hambleton & Powell 1983), judges are asked to rate an item for the probability that students would get the item correct if they were minimally competent. There is only one judgment per item, i.e., for the minimally competent examinee, and for whether the student would get the item correct. In the NAGB procedure, both of these elements were modified to meet the Board's policy.

In setting achievement levels, every item was being rated three times, once for BASIC, again for PROFICIENT, and finally for ADVANCED. Moreover, the judgment was not based on the probability of whether the examinee would get the item correct, but rather, whether the examinee should get the item correct; that is, if the examiner were BASIC, PROFICIENT, or ADVANCED.

In the Vermont meeting, the judges received the complete item pool, including the higher order thinking skills (HOTS) and estimation (EST) items, on which to make judgments. It was deemed advisable to provide the complete item sets because at that time it was unknown whether or not the HOTS and EST items would scale properly. If the operations contractor was able to scale this component of the full item pool, then these items could be included in reporting the achievement levels. In the final analysis, these items were not capable of being scaled with the remaining items and were reported separately and without regard to the achievement levels. The following table shows the distribution of items in the pool:

Grade	Core	HOTS	EST	Total
4	109	14	20	143
8	135	8	46	189
12	143	13	46	202

There were distinct differences between the core item pool and the special study blocks. First, the special study blocks were administered under different conditions than the core, i.e., using a paced-tape. Secondly, the special study blocks were not included in the administration of the Trial State Assessment (TSA) because of limited resources; in other words, the TSA included only blocks 3 to 9. Blocks 10 to 12 were administered only to a subsample of the national sample. Even if these items had been able to be linked to the mathematics composite scores[1], the advisability of including them in the achievement level-setting process was certainly questionable since they had not been administered as part of the TSA.

---

[1] An internal ETS memorandum documenting technical reasons for not linking the HOTS and EST items to the math composite is dated September 27, 1990 from G. Johnson et al to S. Koffler and I. Mullis.

The item-rating task was similar for both multiple-choice and production (open-ended) items. The judges were instructed to review the item, work it out, check their answers against the key provided, and then to make a judgment about the number of examinees out of a group of 100 marginally BASIC, PROFICIENT, or ADVANCED who should get the item correct. Item ratings were summed across all items to calculate the three achievement levels and then averaged across all judges to obtain achievement levels from the total group.

3.6 Content Descriptions of the Achievement Levels

The value of setting achievement levels is not so much in the achievement levels per se, but in the competencies that examinees at those achievement levels can demonstrate. In order to describe the mathematical skills and behaviors of BASIC, PROFICIENT, and ADVANCED students, it was necessary to try to employ the judges' ratings of the items to construct these content descriptions. Essentially, this involved looking at an individual item's ratings and identifying those items whose probability ratings were substantially higher for PROFICIENT than for BASIC, and higher for ADVANCED than for PROFICIENT. Items whose ratings were judged to be more PROFICIENT than BASIC, or more ADVANCED than PROFICIENT were then clustered, and content patterns examined.

To illustrate this process, the judges' ratings on five items are listed below. In examining the judges' ratings the 80/50 rule was used. Items that were judged to be about 50 percent or less for the BASIC level and about 80 percent or more for the PROFICIENT level were selected as possible representatives of the PROFICIENT level; those judged to be about 50 percent or less for the BASIC and or PROFICIENT levels, but 80 percent or more for the ADVANCED level were selected as possible representatives of the ADVANCED level; items that were judged at or above 80 percent for the BASIC level were selected as representative items of the BASIC level.

In the sample below, items 1 and 4 would be BASIC items; item 5 PROFICIENT; and items 2 and 3 ADVANCED.

Item No.	BASIC	PROFICIENT	ADVANCED
1	0.82%	0.91%	0.99%
2	0.23	0.47	0.81
3	0.37	0.53	0.86
4	0.78	0.89	0.92
5	0.48	0.76	0.87

Mathematical definitions were then developed from these content clusters by a subgroup of the participants in the Vermont meeting. Eleven mathematics and curriculum experts were selected to develop the definitions based on the round four judges' ratings. They also selected from the released item pool those items that best exemplified the content descriptions they had developed.

Finally, the subgroup developing the definitions verified the sample items using the item characteristic curves (ICCs), which have been available since November 1990. For each sample item identified, panel members estimated from the ICCs the probability of an examinee answering the item correctly at the achievement level (projected onto the NAEP scale) for the particular level which the item was to represent. Again, a probability of 0.80 was used to confirm the appropriateness of the sample item for a given level.

4. Analysis of Achievement Level Ratings: Phase 1

4.1 Introduction

Tables summarizing the analyses of the data collected during the Vermont/Washington phase of the project are contained in appendices F and G. The 39 tables in appendix F displaying data from Vermont and Washington replace all previous drafts of tables that have been circulated. Changes that have been made from earlier drafts are minor and do not affect any substantive interpretations or criticisms that have been directed at the results.

4.2 Overview of Results

Tables 1 to 3 provide a summary of the achievement levels at all five rounds of the process for grades 4, 8, and 12, respectively. Readers may refer to tables 15 to 17 for the individual judges' achievement levels. A few points about tables 1 to 3 can be highlighted:

1. Except at grade 8, the number of judges dropped substantially between rounds one and three (Vermont meeting) and rounds four and five (Washington meeting). This fact must be kept in mind when interpreting the statistics across the rounds in nearly all of the analyses.
2. At grade 4, the BASIC achievement level remained nearly constant over the five rounds. The PROFICIENT and ADVANCED levels moved up 4 to 5 percent. In all cases but one, the variability of ratings decreased from the first to the fifth round, and at the fifth round, variability across judges appeared to be quite low.
3. At grade 8, there was a definite pattern for the achievement levels to drop from 3 percent (ADVANCED) to 6 percent (BASIC and PROFICIENT) over the five sets of ratings, with the biggest drop occurring between the fourth and fifth

rounds. It was between the fourth and fifth rounds that the consistency and coherence of the achievement levels was considered by the judges, and the grade 8 judges were made aware that their ratings appeared to be out of line with the ratings at grades 4 and 12. This point will be expanded upon below. The variability among the grade 8 judges was considerably higher than at the other two grade levels. Some convergence can be seen in that the standard deviations of the ratings dropped about a third between the first and last rounds. At the other grade levels, the decrease in variability over the rounds was far greater.

4. At grade 12, changes over the five sets of ratings were mixed. BASIC went up 3 percent; PROFICIENT and ADVANCED dropped 3 to 4 percent. Variability among the judges generally decreased over the five sets of ratings (there were two exceptions and both occurred between rounds three and four suggesting that perhaps the samples were different). Clearly though, the discussions between rounds four and five substantially influenced the results--PROFICIENT and ADVANCED levels dropped 3 to 5 percent and variability among the judges also dropped substantially too.

Independent of any of our analyses, ETS decided to report performance on the estimation and higher-order thinking skills items differently from the remaining cognitive items of the mathematics assessment due to problems in referencing these items to the NAEP reporting scales. It seemed advisable therefore to recompute the statistics in tables 1 to 3 using the reduced set of test items (about 25 percent of the items at each grade level were dropped). Tables 4 to 6 provide the same information as tables 1, 2, and 3 based on the reduced item pools. Of course at round five, recalculations of achievement levels could not be done because item ratings were not available. In addition, since in subsequent analyses we had determined that several of the distributions of judges' ratings were skewed, tables 4, 5, and 6 contain both the median and the mean ratings. The complete set of individual judges' achievement levels across the five sets of ratings in the reduced item pool are contained in tables 27 to 29.

A review of tables 4, 5, and 6 versus tables 1, 2, and 3 led to the following observations:

1. The revised (reduced item pool) achievement levels were up slightly at grade 4 (apparently the deleted items were judged to be relatively harder than those that remain in the pool), unchanged at grade 8, and lower at grade 12 (apparently the deleted items were judged to be relatively easier).
2. On the basis of the final round of ratings, it appeared that several of the distributions of judges' ratings were skewed: grade 4 PROFICIENT, grade 8 BASIC, grade 12 PROFICIENT. This point was addressed during the final stages of the analyses (see appendix G).

Tables 7, 8, and 9 provide information about the ratings of the four groups at each grade level on round three (following discussion among judges in each group). Comparisons of achievement levels (means) across the four groups within each grade level could be thought of as checking the consistency of results across different groups of judges. Such a comparison of means could provide a basis for estimating the standard errors of the achievement levels, albeit on samples one-fourth the size of the total group, at a stage prior to the final ratings. The comparison is meaningful only as an estimate of standard error when the groups can be considered to be drawn at random from the population of judges of interest. The results seem to indicate:

1. At the grade 4 level, the range of achievement levels (means) across the four groups at round three was 16 percent BASIC, 13 percent PROFICIENT, and 7 percent ADVANCED.

2. At the grade 8 level, the range of achievement levels (means) across the four groups at round three was 27 percent, 20 percent, and 9 percent, for the three levels, respectively.

3. At the grade 12 level, the range of achievement levels (means) across the four groups at round three was 20 percent, 10 percent, and 5 percent for the three levels, respectively.

These results definitely show more variability than would seem desirable; however, the achievement level-setting process was never intended to stop at round three. Also, the equivalence of the groups at the time of formation was never established either. Some of the group differences may have existed before the work actually started.

To examine the "equivalence of groups" hypothesis the reader is referred to tables 37, 38, and 39. It is clear that the groups were not equivalent initially since the means varied widely and the standard deviations were quite large for round one. Even though equivalency was desirable, and even ostensibly present in the assignment of individuals to the different groups, in fact, this simply was not the case. Individuals within the groups were interpreting the generic definitions differently perhaps, and came with their own sense of what examinees should know and be able to do. This is clearly reflected in the round one data. Therefore, readers must interpret with caution the round three data and its level of variability.

One of the unique and useful features of the 1990 NAEP mathematics assessment was the presence of 32 items common to the three grade level assessments, 27 items common to grades 4 and 8, and 78 items common to grades 8 and 12. Tables 10, 11, and 12 provide information on the locations of the common items in the item bank booklet at each grade level, and the judges' third round ratings on each common item. A review of the statistics in these tables revealed that the grade 8 item ratings appeared to be inconsistent. On common items, the grade

8 judges set higher achievement levels than the grade 12 judges. Table 13 summarizes the actual 1990 student performance on the common items. (To simplify the analyses, a 50 percent random sample of items common to grades 8 and 12 was used.) The patterns are clear:

1. Performance on items increased with the amount of schooling.
2. Items which were common to grades 4, 8, and 12 tended to be relatively hard for grade 4 (.42 compared with .48 for the total grade 4 pool) and relatively easy for grade 12 (.76 compared with .55 for the total grade 12 pool). Similar patterns were noted for items common to grades 4 and 8 and grades 8 and 12.

Table 14 highlights the problem revealed by our analyses of tables 10 to 12. Using only the common items to set achievement levels would result in a higher achievement level at grade 8 BASIC than grade 12 BASIC, and near identical achievement levels for PROFICIENT and ADVANCED. In fact, the average student showed an actual increase in performance of 14 percent on the common items at grades 8 and 12. When confronted with these results, after round four, the grade 8 judges lowered their BASIC achievement level and the grade 12 judges increased their BASIC achievement level. During the discussion of the reasons for why the judges had rated the items as they did, it became clear that there were some substantive reasons to account for their judgments. It was argued by some judges that the content of the common items was such that they reflected the content that was generally covered in the seventh-eighth grade sequence, and not in the high school mathematics course work (if students were even enrolled in such courses). Therefore, it was more likely that eighth graders would perform better on these common items than would twelfth graders, who could be as much as 4 years removed from any formal instruction in these areas. Though the reversals in the third round ratings were troublesome (noting the amount of changes in the achievement levels between the third round and the final (adjusted) achievement levels), it would seem likely that the majority of reversals would have been eliminated. For example, the grade 8 BASIC level was lowered by 11 percent and the grade 12 BASIC level was lowered by 1 percent for a difference of 10 percent. At round three, the grade 8 BASIC level exceeded the grade 12 BASIC level by 5 percent using the common items only. The subsequent ratings and adjustments would have reversed the situation at round three, and the grade 12 BASIC level would exceed the grade 8 BASIC level by 5 percent. Still, the difference in achievement levels for four years of school seems small, given the potential room for growth (note that the adjusted grade 12 BASIC level on the common items was .72).

Tables 18 and 19 summarize the achievement levels based on the total pool of items. These numbers were shared with the judges at the completion of the process. It was only later that the statistics in tables 4, 5, and 6 were calculated, and the final achievement levels were based upon the reduced item pools. The variability of achievement levels at grade 8 BASIC remained very high, while at other levels and grades the variability seemed a little higher than might be desirable, though it is important to keep in mind that NAGB had intentionally chosen a diverse pool of judges, including 30 percent from outside the field of education.

Table 20 contains the confidence ratings associated with the final achievement levels reported in Tables 18 and 19. Of the 114 ratings provided by the 38 judges, 110 were ratings of "confident" or "very confident" and 4 were of ratings "somewhat confident" (two of the four were at the grade 12 BASIC level).

4.3 Attrition Prior to the Washington Meeting

One of the troublesome aspects of the achievement level-setting process was that 24 of the 63 judges were unable to return to Washington for the second meeting on September 29-30,

1990. Tables 21 to 23 summarize the statistical data on the groups of judges who returned to Washington versus those who did not.

It should be noted that all 63 of the original judges were formally invited to participate in the followup meeting. Letters were sent to all the judges, explaining the need for a second meeting and indicating what tasks would be accomplished at that meeting, which was held over a weekend to encourage the participation of teachers and others who might already have weekday commitments. However, these days (September 29-30) were also religious holidays for some of the participants, which accounted for about 50 percent of those who did not return. A telephone survey of many of those who indicated they would not attend showed that prior commitments accounted for the remaining 50 percent.

It was unfortunate that this particular weekend was selected. However, since the federal government did not have a budget as of October 1, 1990, it was considered in the best interest of the project to try to have the meeting before any fiscal disruption took place.

The main findings from tables 21 to 23 show that:

1. Nearly all of the grade 8 judges returned. No concerns were raised about the two missing judges. At grades 4 and 12, the loss of judges was about 50 percent and concerns were raised.
2. A disproportionate number of noneducators were unable to attend the Washington meeting.
3. At grade 4, the nonreturning judges had set higher achievement levels than those who did return to Washington.

A complete analysis of these data is contained in appendix G. The final outcome was that while adjustments seemed warranted, especially at grade 4, insufficient evidence was available to decide on either the nature or the amount of the adjustment. Therefore, no adjustments were made to correct for the changing character of the pool of judges who participated at the Washington meeting.

4.4 Explanation of the Adjustments in Tables 24, 25, and 26

The statistics in Tables 24 to 26 were used by 12 judges in preparing skill descriptions of the marginally BASIC, PROFICIENT, and ADVANCED students. The numbers in Tables 24 to 26 are the (adjusted) averages of the total group of judges' achievement levels at the item level from round four. Of course, these 12 judges should have used the item statistics based on the final (fifth) round of ratings, but these ratings were not provided at the item level. Therefore, the item ratings at the fourth round were used to reflect the final item ratings, but they were adjusted to highlight changes in the overall achievement levels between the fourth and final ratings. The adjustments based upon mean achievement levels in Tables 4, 5, and 6 are shown below:

Level	4th Round	Final Round	Adjustment
	Grade 4		
Basic	49.4%	50.5%	+1%
Proficient	76.5	77.3	+1
Advanced	89.6	90.2	+1
	Grade 8		
Basic	68.9	64.1	-5
Proficient	85.1	81.3	-4
Advanced	93.9	91.8	-3
	Grade 12		
Basic	54.4	56.4	+2
Proficient	81.1	78.0	-3
Advanced	93.4	90.8	-3

Items in tables 24 to 26 without achievement levels were those items that were deleted because they measured EST or HOTS.

4.5 Achievement Levels for Content Categories and Abilities

Tables 30 and 31 highlight the achievement levels at each grade level for the five content categories (table 30) and mathematics abilities (table 31). It is not clear what pattern of achievement levels would most reflect the validity (or invalidity) of the achievement levels. Certainly there is evidence of variability in achievement levels across content categories which might also be expected. Also, achievement levels tended to be higher in the area of numbers and operations than in the other areas which might also be expected. This pattern is fairly clear at grades 4 and 12, but not at grade 8.

One might reasonably hypothesize achievement levels to be lower for problemsolving than for conceptual understanding, which they were by about 10 percent at BASIC, 6 percent at PROFICIENT, and 3 percent at ADVANCED. An analysis of the actual item p-values would probably provide a basis for interpreting the meaningfulness of the achievement levels and their variability. But even the meaningfulness of this analysis is questionable because it is quite possible that valid achievement levels would not follow the same pattern as the actual p-values. Finally, we note that because of the way items are selected (easy, middle difficulty, and hard items within each of the 15 combinations of content and process levels), it is probably impossible to meaningfully hypothesize the valid arrangement of achievement levels in the content categories and ability categories.

4.6 Item Appropriateness Ratings

One potential problem that arose during the Vermont meeting was that a number of the judges questioned the appropriateness of an unspecified number of test items. Judges reacted in different ways. Some judges were able to put their personal views aside and continue with the item rating process. Other judges indicated that they lowered their ratings arguing that these items were less appropriate and therefore lower expectations of performance were reasonable. It was unknown how many judges questioned the appropriateness of the NAEP items, or how they may have been affected.

When the opportunity was there to conduct the Washington meeting, we made the decision to obtain item appropriateness ratings (low, median, or high) from the judges. Tables 32, 33, and 34 provide the descriptive statistics on the item appropriateness ratings for grades 4, 8, and 12, respectively. A summary of the overall results appear in table 35. The results differed substantially across grades. At grade 4, item appropriateness ratings appeared to be very high. At grade 8, the results showed considerably lower item appropriateness ratings. At grade 12, the results were between grade 4 and 8.

4.7 Correlations Between Expected and Actual Item Difficulty Values

One criticism directed by the Technical Review Panel (TRP), a technical group contracted to conduct validity studies for NAEP, at the (third round) achievement levels was the relatively high correlations between the item ratings and the actual item p-values. The argument was that the validity of the resulting achievement levels was lowered because of the critical role of the empirical data at rounds two and three. At the time of their analysis, the TRP did not have access to information that could be used to compute the correlations for all three rounds. Table 36 provides the complete set of correlations. A comparison of the correlations shows that even at round one, perceptions of item difficulty were a prominent factor in the ratings process. The correlations ranged from .57 to .79. At grade 4, the correlations were substantially lower.

5. Conclusions and Recommendations: Phase 1

5.1 Adjustments to the Phase 1 Achievement Levels

Based on all the evidence collected, it was clear that there were concerns about the recommended levels based on the Vermont/Washington meetings. There were several reasons for this. Readers are referred to appendix G for a detailed analysis of those concerns that had implications for adjusting the achievement levels.

The elimination of the HOTS and EST items (which had been decided late in the process based on empirical evidence of the lack-of-fit of these items on the composite scale) necessitated an adjustment in the data collected in Vermont. The adjustment of the data set by removing the ratings of the HOTS and EST items from the judges estimates was straightforward enough. However, an important question raised by the evaluation team needed an answer: Was there a contextual problem here? In other words, if the judges in Vermont had never seen the HOTS and EST items, would they have judged the remainder of the item pool differently? This was a moot question at this point, because in fact the judges had seen the HOTS and EST items, and rated them three times.

Second, there was the issue of missing data. Sixty-three judges participated in the Vermont meeting, while only 39 participated in the Washington meeting -- a 40 percent shortfall. This problem was somewhat more complicated to deal with.

One question to be answered was: Did the missing judges tend to set higher (or lower) achievement levels than those who attended the Washington meeting? Estimates based on the earlier data collected in Vermont tended to show that the missing judges did indeed set somewhat higher achievement levels than did those who attended the second meeting in Washington (particularly at grade 4). An analysis of the data by educator/noneducator subgroups also showed that noneducators tended to set higher achievement levels by 4 to 6 percent; and many of them were missing from the Washington meeting.

Finally, because of the resulting skewed distributions of the judges' ratings, it seemed to be advisable to use the median of ratings instead of the means of ratings in setting the achievement levels.

At this point in the process, the TACSS discussed each of these issues and came to the following recommendations for the Board:

1. Adjust round five data at all grades to account for the reduced item pool (elimination of HOTS and EST items).

2. Adjust round five also at all grades to account for skewness in the ratings by using medians instead of means.

3. Do not adjust the ratings to address the missing judges at the Washington meeting.

On the last point, adjustments at grade 4 seemed necessary, but there did not seem to be a defensible basis on which to make adjustments. For one, the sample sizes were too small to estimate any adjustments reliably.

5.2 External Evaluations of the Level-Setting Process

There were several external evaluations being conducted throughout most of phase 1. The Board itself had contracted with the evaluation team from Western Michigan University. In addition, the National Center for Education Statistics, under whose auspices the NAEP program is implemented, directed the Technical Review Panel (TRP), a technical group contracted to conduct validity studies for NAEP, to conduct a meta-analysis of the data collected in the process. Further, various stakeholder groups such as the Council of Chief State School Officers (CCSSO)--the agency which conducted the national consensus process to develop the 1990 mathematics framework for the Board and which holds a stakeholder interest in the achievement level-setting process since 37 states, the District of Columbia and 2 territories were participants in the 1990 Trial State Assessment--were keeping close watch on what was happening. The Education Information Advisory Committee (EIAC) of the CCSSO provided some very positive recommendations to the Board throughout the process.

Each of these groups, and others not mentioned, expressed serious concerns about the achievement levels resulting from the Vermont/Washington meetings. By the beginning of January 1991, the Board was faced with a dilemma. Should it abandon the work done so far, and start all over again, or should it continue on and try to validate the levels which it now had?

5.3 A Summary of the Problems

On the surface, the achievement level-setting task had seemed straightforward. After all, most advisers and consultants who were involved had read the relevant standard-setting literature and had conducted a number of these standard-setting studies in the past. Not surprisingly, the popular Angoff standard-setting method (Angoff, 1971) was selected; judges would be identified and trained, and then they would complete their ratings, and the levels would be determined. Along the way, consultants would be involved who would keep the project on an acceptable technical course.

Unfortunately, problems in implementation did arise. Perhaps some of the problems should have been detected; others could not have been foreseen. For example:

a. From the beginning, there was always pressure to move more quickly than might have been desirable. Production schedules were already set at ETS for NAEP data analysis. This project needed to meet those production schedules, or the desired reports could not be produced.

b. The number of participating judges was large and diverse--70 percent educators, 30 percent noneducators. These individuals were important persons in their own area of expertise, quite articulate, came to the process with many questions, and, in some cases, with their own agendas. Each judge wanted to do the very best possible job, but available time was lost in responding to the many issues and questions raised by the judges.

c. The quality and appropriateness of the item pool came under attack from some judges. Without passing judgment on the validity of the criticisms, for many judges, the task became more complex. They were simultaneously trying to balance item difficulty with item appropriateness and even item quality. For example, if the item is easy but inappropriate, what rating should it be given?

d. Judges were asked to specify how examinees should perform on items. This is a considerably more difficult task than asking judges how students would perform.

e. Judges were asked to provide three ratings for each item. They were asked to specify how marginally BASIC, PROFICIENT, and ADVANCED students would perform. Again, this task is considerably more difficult and time consuming than setting one level, as is more customary (Busch & Jaeger 1990).

f. Judges were working at one of the three grade levels, but Board policy dictated consistency and coherence for the final achievement levels across grades. For example, it would make little sense, and would threaten the validity of the process, if the achievement level for the BASIC student at grade 8 exceeded the achievement level for the BASIC student at grade 12 (after corrections are made for test difficulty at each grade level).

g. The definitions of BASIC, PROFICIENT, and ADVANCED were specified by Board policy, but these were generic definitions that would apply to many subject areas. The result was that the definitions proved difficult to work with at the operational level.

h. Test lengths at each grade level were large, exceeding 100 items, and at grade 12, over 200 items. This factor contributed to making the task more difficult as well.

i. The actual ratings were carried out in a "fish bowl." The NAGB staff, ETS staff, NCES staff, NAGB Board members, the evaluation team, the Trial State Assessment evaluators, the training staff, and even a news reporter, were present in the room where the process was taking place.

In fact, despite some of the difficult hurdles to overcome, and because of the very hard work of the judges, the full process as scheduled, with some midcourse corrections, was completed, after more than 1,600 hours of volunteered time from the judges.

5.4 Recommendations

Many of the criticisms directed at the process by the Board's evaluators, the TRP, the TSA evaluators, the stakeholder groups, and even the judges themselves appeared to be correctable, or, at the very least, could be ameliorated, if the process was conducted again for the purpose of validating the levels. The Board, therefore, decided in February 1991, after conducting a public hearing on the Vermont/Washington levels, to validate those levels through a replication/validation study. This study would be conducted in the late winter and early spring, and the results would be reviewed and discussed at the May meeting of the Board.

6. The Replication/Validation Study: Phase 2

6.1 Introduction

The work on the first effort to set achievement levels in mathematics has shown both the importance and the complexity of the task. After more than a year, additional work was still required before the Board could reach a decision regarding the 1990 mathematics achievement levels. Enough work had been completed up to this point on the initial effort to set mathematics achievement levels to allow individuals and groups to comment on both the process and the product. Several extensive evaluations or secondary analyses were now completed that contributed to a fuller understanding of the proposed levels and that provided both technical and policy commentary on the levels and how they were derived. These commentaries raised issues about the levels that needed to be addressed as the Board moved ahead with its plan to report the 1990 NAEP mathematics results and to develop achievement levels for 1992 and beyond.

The Board, therefore, consistent with its role as the policymaking body for NAEP, and taking the advice of many thoughtful groups and individuals, decided to conduct a validation study of the achievement levels before reaching any final decision. The validation process consisted of a series of activities designed to provide evidence of validity for the achievement levels. The five major components of the process are described below.

6.2 Replication/Validation Study

The plan described here was approved on February 12, 1991, by the two Board committees responsible for monitoring the achievement levels process. It was developed by the NAGB staff in consultation with the Ad Hoc Advisory Committee on Achievement Levels Validation. Participating in the Ad Hoc Committee meeting, and in the subsequent review of materials, were Peter Behuniak, Connecticut Department of Education; Thomas Fisher, Florida Department of Education; Ronald K. Hambleton, University of Massachusetts; Marilyn Hala, National Council of Teachers of Mathematics; Anne Hess Lockwood, National Computer Systems; Tej Pandey, California Department of Education; Edward Roeber, Michigan Department of Education; and Ramsay Selden, Council of Chief State School Officers.

The Ad Hoc Committee reviewed the initially recommended levels, the descriptions and sample items; a profile of the initial achievement level-setting panel; the results of a survey of the panelists' approval of the levels; the CCSSO board of directors' statement; selected state responses to the levels; written technical documentation about phase 1; the Western Michigan University interim evaluation report; testimony from the public hearing on January 8, 1991; an executive summary of the Technical Review Panel report; and various media articles.

Based on the evidence at hand, the Ad Hoc Committee concurred with the staff proposal to conduct a validation study, suggesting that some attention be given to replicating the original process as much as possible. The following briefly describes each task of the plan.

Task 1: Technical Report

It was mentioned earlier that the Board undertook this initiative more than 14 months ago. During this period, many aspects of the project were completed (materials were produced for meetings, documents developed as a result of meetings, and many individuals and groups were involved.) While this documentation existed, it had not yet been systematically collected and presented in the form of a technical report. This was required if the process was to be understood and accepted.

Therefore, this technical report was prepared as part of the validation study. It addresses the technical aspects of the process, as well as the Board policies implemented through various technical decisions.

Task 2: Executive Summary

As important as the technical report may be, a shorter, less technical summary was also important. The work of the Board and the product they were considering needed to be accessible, understandable, and useful to a wide audience of stakeholders, interest groups, and publics, including legislators, federal, state, and local policymakers, the business and industrial communities, and most especially teachers, parents, and students. Therefore, a short, focused summary of the achievement levels process, including the next steps to be taken in the validation process, was prepared to respond to the needs of this larger audience. The substance of this summary is included in this report as Chapter 1, initially prepared by Larry Feinburg, NAGB Assistant Director for Reporting and Dissemination, and further edited by the authors of this report.

Task 3: Site Validations

The centerpiece of the validation effort consisted of four state meetings in various regions of the country designed to collect structured feedback on the proposed achievement levels.

Since NAEP collects data from students representing each region of the country, four meetings were held in March--one each in the Northeast, South, Midwest, and West. Four state departments of education offered to assist the Board in conducting these meetings, including California, Connecticut, Florida, and Michigan. The details of selecting and training judges and the item rating tasks are described in subsequent sections of this chapter.

Task 4: Final Review by Math Panel

The original study plan called for reconvening a subgroup of the 63-member Vermont panel to review the data collected in the validation effort. If the results of the validation produced achievement levels that were substantially the same as those initially recommended from the Vermont/Washington meetings, then there would be a need for only modest revisions. Alternately, if the results of the validation produced results that were significantly different from those produced in the original process, the work of this subgroup would be to develop some recommended options from which the Board could make its final decision.

In actuality, because of the pressures of time, three members of the Vermont groups--John Dossey, professor of mathematics at Illinois State University; Mary Lindquist, Columbus College in Columbus, Georgia; president-elect of the National Council of Teachers of Mathematics Steve Lienwand, mathematics consultant with the Connecticut Department of Education; and Martha Bacca (not a member of the Vermont panel) from Phoenix--reviewed the validation data, developed the definitions, and recommended selected released items for the achievement levels.

Task 5: Response to Evaluations

While this technical report and executive summary no doubt will address some of the issues raised through the Western Michigan evaluation, the Technical Review Panel's secondary analyses, and the National Academy's Trial State Assessment evaluation, there was no mechanism for correcting factual errors, or for presenting competing explanations of the data. A formal rejoinder was required in the Replication/Validation plan to "set the record straight" and to present alternative hypotheses or interpretations of the findings. Some additional analyses were required, and some additional data collection from the panelists was considered. Responding to criticisms in a reasoned way, and from a data-based posture, is an essential aspect of the validation process. Tasks 1, 2, and 3 alone would not answer all the questions raised in these documents. Task 5 was viewed by the Board as critical since this is a trial program, and debate and discussions of both the methods of achievement level setting and the results are important for technical and policy reasons. Task 5, however, is an ongoing activity. This report is a first step. The authors hope that future discussions through publications and paper-presentations will continue to illumine the debate.

6.3 Selection of Judges

Approximately forty-eight mathematics teachers and twelve noneducators were invited to participate in one-day sessions. The criteria for teacher participation were: (1) teachers must currently provide direct instructional services in mathematics to students in grades 4, 8, or 12, and must represent teachers of students with varying ability levels; (2) as a whole, the regional group must be representative on the basis of gender and ethnicity; (3) as a whole, the regional group must include both novice and experienced teachers, and must be drawn from urban, suburban, and rural communities of varying sizes.

The criteria for selection of noneducators was the same as the criteria that was used to identify participants for the original panel--that is, leaders of business and industry, professional groups, parents, individuals who have shown an interest in education, as well as persons who have initiated or implemented school-business partnerships, were all eligible candidates. Naturally, those selected should contribute to the overall representativeness of the group in terms of gender and ethnicity.

The state education department representatives assisted in identifying teachers and noneducators in their state or region who collectively met these criteria.

6.4 Training of Judges

The one-day session included a modified training activity for participants, an independent rating of a sample of items, an opportunity for participants to judge the proposed achievement levels against their own ratings, and to comment on the proposed achievement levels, descriptions, and sample items. Written, structured feedback was solicited from each participant with no attempt to reach consensus. This information was synthesized for and presented to the Board as they made their final decision.

A scripted videotape was prepared so that all four presentations were standardized, and participants would not be biased by the presenter in their approach to the task. This approach also ensured consistency in training and group preparation. The tape was divided into four segments: (1) introduction to the process; (2) initial training and preparation of the group; (3) calculation of ratings and comparison of these ratings with proposed achievement levels; and (4) collection of structured feedback. The tape systematically led the group through the packet of materials distributed at the meeting. Mary Lyn Bourque, NAGB Assistant Director for Psychometrics, was responsible for coordinating the meeting, ensuring a standardized approach, and answering questions from the participants.

6.5 Item Rating Tasks

All procedures were field tested locally before any meetings were conducted so that the scripts could be refined and finalized, and timing of the tasks (which was such a problem in earlier meetings) could be properly scheduled.

Each participant was asked to provide **one set** of ratings for a marginally BASIC, PROFICIENT, and ADVANCED group of students on a sample of items. Since item samples were already part of the NAEP BIB spiral design, actual NAEP item booklets were used by the participants. They also had the appropriate manipulables such as calculators, protractors, and rulers. Approximately 70 participants across all sites rated one of seven test booklets at each grade level, which yielded about seven ratings per item per site, or 29 ratings per item across all four meetings. In addition to timesaving, this arrangement met the need for ensuring better item security by not divulging the entire item pool to each participant.

After providing an independent rating of the item samples, each participant was instructed in how to estimate their sample achievement levels. They were also given the achievement levels of the original panel and other relevant data and then asked to critique the achievement levels in the light of their own professional judgment. In addition, participants were asked to provide commentary on the proposed descriptions and the sample items associated with the levels. This commentary was collected using feedback protocols specifically structured to probe the issues (e.g., whether there was sufficient justification for an ADVANCED level given the content of the assessment).

6.6 Description of the Levels

On the pages that follow the complete descriptions developed through the validation study are displayed, as well as the corresponding achievement levels, and the sample items for each level.

6.7 Summary

While the validation procedures may appear at first glance to be a short-term process, the work of validation is a continuing one which is expected to proceed well beyond the five tasks described earlier. For example, one of the Board's initial goals in exploring achievement levels as a reporting mechanism was to "improve the form and use of NAEP results." Therefore, if the results of the 1990 mathematics assessment are reported in terms of the achievement levels, it would be advisable for the Board to gather evidence on the utility of the levels to users of NAEP data. The utility and understandability for policymakers, which can only be obtained after the results are released on September 30, is an important component of determining the intrinsic value of setting achievement levels on any assessment, especially NAEP.

In addition, at the time of this writing, the Board is expected to set achievement levels again in 1992 in mathematics, and in reading and writing as well. But it is noted the levels set for 1990 are trial levels, and should not be used as benchmarks for measuring progress in the nineties unless there is ample evidence that the achievement levels are reliable and valid for the use to which they will be put.

Figure 3-- Mathematics Proficiency Corresponding to Each Achievement Level, By Grade: For 1990 NAEP Mathematics Assessment

GRADE ACHIEVEMENT LEVEL	PERCENT CORRECT*	MATHEMATICS PROFICIENCY*
Grade 4		
Basic	45	207
Proficient	68	245
Advanced	87	283
Grade 8		
Basic	48	255
Proficient	72	295
Advanced	89	336
Grade 12		
Basic	47	282
Proficient	73	330
Advanced	88	358

*The percent correct is the proportion of items that students should answer correctly in order to reach each level. The percent correct scores were then transformed to the proficiencies on the new NAEP mathematics scale used to produce the statistical summaries.

Exhibit 1: Levels of Mathematics Achievement for Grade 4

(283) ADVANCED: Superior Performance

Fourth-grade students who are performing at the advanced level should be able to demonstrate flexibility in solving problems and relating knowledge to new situations. They should be able to use whole numbers to analyze more complex problems. Their understanding of fractions and decimals should extend to a number of representations. Students at this level should determine when estimation or calculator use is an appropriate solution to a problem, as well as read and interpret complex graphs. Advanced fourth-grade students should also be able to use measuring instruments in non-routine ways. These students should be able to solve simple problems involving geometric concepts and chance.

(245) PROFICIENT: Solid Academic Performance

Fourth-grade students who are performing at the proficient level should have an understanding of numbers and their application to situations from students' daily lives. The proficient student should be able to solve a wide variety of mathematical problems; use patterns and relationships to analyze mathematical situations; relate physical materials, pictures, and diagrams to mathematical ideas; and find and use relevant information in problem solving. Fourth-grade proficient students should understand numbers and concepts of place value and have an understanding of whole number operations, as well as a facility with whole number computation. For example, students should be able to solve problems with a calculator and have the ability to use estimation skills to solve problems. Proficient fourth-grade students should understand and use measurement concepts such as length; be able to collect, interpret, and display data; and use simple measurement instruments.

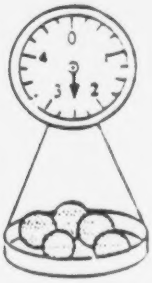
(207) BASIC: Partial Mastery of Knowledge and Skills

Fourth-grade students who are performing at the basic level should be able to solve routine one-step problems involving whole numbers with and without the use of a calculator. They should also be able to use physical materials and pictures to help them understand and explain mathematical concepts and procedures. Students at this level are beginning to develop estimation skills in measurement and number situations and should understand the meaning of whole number operations. For example, students performing at the basic level should be able to link the meaning of multiplication with the symbols needed to represent it. These students are also beginning to develop concepts related to fractions and read simple measurement instruments. Basic fourth-grade students should also be able to identify simple geometric figures and extend simple patterns involving geometric figures. These students should be able to read and use information from simple bar graphs.

Grade 4 Basic: Example 1

Grade 4: 76% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
73%	94%	98%



The scale shown above measures weight in pounds. What is the total weight of the oranges in the picture?

[✓ A] $2\frac{1}{2}$ pounds

B $3\frac{1}{2}$ pounds

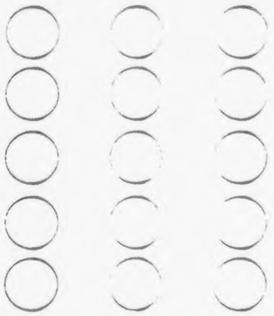
C 5 pounds

D 10 pounds

Grade 4 Basic: Example 2

Grade 4: 80% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
79%	95%	100%



Write a multiplication sentence to find the number of circles.

5 × 3 = 15

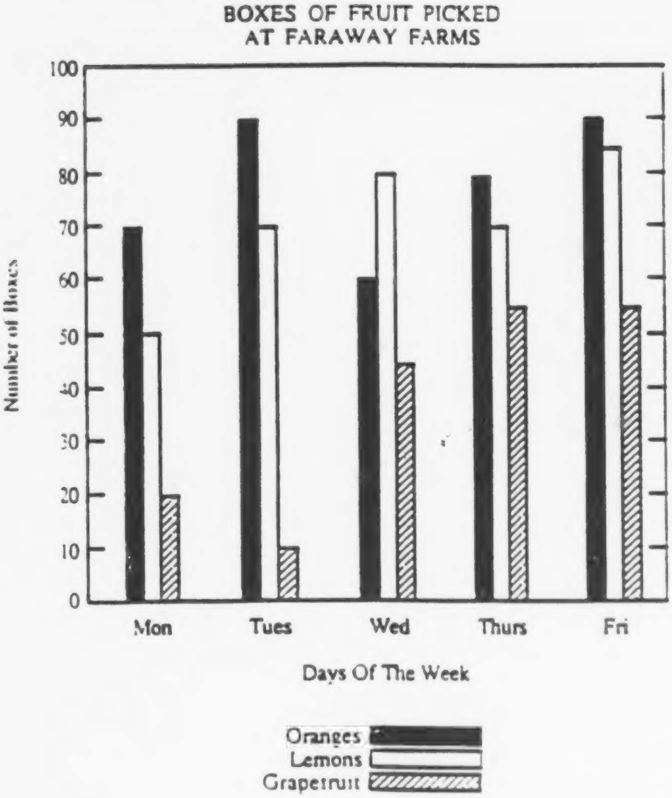
Grade 4 Basic: Example 3

Grade 4: 80% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
79%	90%	98%

Grade 8: 89% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
88%	94%	94%



How many boxes of oranges were picked on Thursday?

A 55

B 60

C 70

[✓ D] 80

E 90

F I don't know.

Grade 4 Proficient: Example 1

Grade 4: 61% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
51%	79%	99%

On a flight from Los Angeles to New York, the cost of a fare was $400. Every seat was sold. What additional information do you need to find the total for all fares?

A None

B The number of employees on the plane

[✓ C] The number of passenger seats on the plane

D The distance from Los Angeles to New York

Did you use the calculator on this question?

○ Yes ○ No

Grade 4 Proficient: Example 2

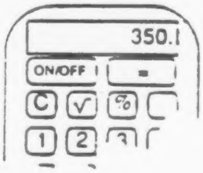
Grade 4: 60% Correct Overall

Percent Correct At Each Achievement Level

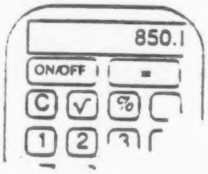
Basic	Proficient	Advanced
54%	75%	84%

The third grade collected more than 850 bottle caps for an art project. The fourth grade collected more than 500 bottle caps. Using her calculator, Maria found the exact total of all the bottle caps collected by both grades. Which calculator could be hers?

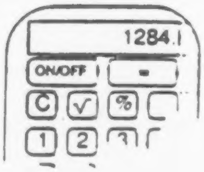
Ⓐ



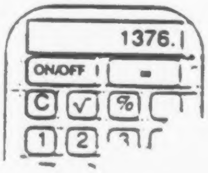
Ⓑ



Ⓒ



[✓ Ⓓ]



Did you use the calculator on this question?

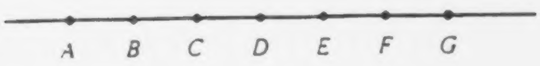
○ Yes ○ No

Grade 4 Proficient: Example 3

Grade 4: 60% Percent Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
54%	84%	97%



In the figure above, points labeled *A* through *G* are spaced evenly along a line. Which of the following distances is the greatest?

A From *A* to *D*

B From *C* to *F*

C From *E* to *G*

[✓ D] From *E* to *A*

Grade 4 Advanced: Example 1

Grade 4: 37% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
34%	38%	64%

Students in Mrs. Johnson's class were asked to tell why $\frac{4}{5}$ is greater than $\frac{2}{3}$. Whose reason is best?

A Kelly said, "Because 4 is greater than 2."

B Keri said, "Because 5 is larger than 3."

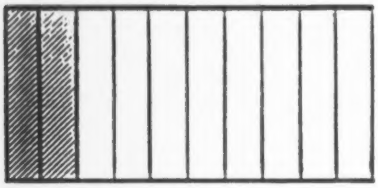
[✓ C] Kim said, "Because $\frac{4}{5}$ is closer than $\frac{2}{3}$ to 1."

D Kevin said, "Because 4 + 5 is more than 2 + 3."

Grade 4 Advanced: Example 2

Grade 4: 61% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
56%	71%	79%



Which decimal represents the shaded part of the figure?

A 0.5

B 0.28

[✓ C] 0.2

D 0.02

Grade 4 Advanced: Example 3

Grade 4: 15% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
6%	28%	72%

The table below shows some number pairs. The following rule was used to find each number in column B.

Rule: Multiply the number in column A by itself and then add 3. Fill in the missing number, using the same rule.

	A	B
Example:	2	7 = (2 × 2) + 3
	3	12
	5	28
	8	67

Did you use the calculator on this question?

○ Yes ○ No

Exhibit 2: Levels of Mathematics Achievement for Grade 8

---

(336) ADVANCED: Superior Performance

Eighth-grade students performing at the advanced level should be able to solve, with and without a calculator, a wide range of practical problems involving percents, proportions, and exponents. These students should have a solid conceptual understanding of the interrelationships among fractions, decimals, and percents and their connections with proportions. Eighth-grade advanced students should also understand and be able to use scale drawings, metric measurements, volume, and accuracy of measurement. These students should be able to solve problems involving elementary concepts of probability, interpret line graphs, and apply basic geometric properties related to triangles and to perpendicular and parallel lines.

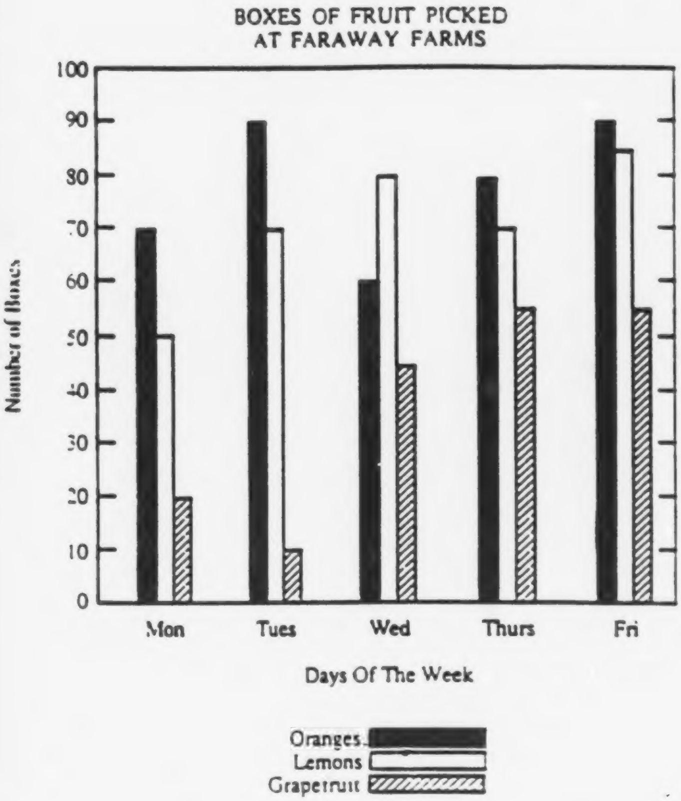
(295) PROFICIENT: Solid Academic Performance

Students at the proficient level should be able, with and without a calculator, to solve problems requiring decimals, fractions, and proportions. They should be able to compute with integers. They should be able to classify geometric figures based on their properties. Proficient eighth-grade students should be able to read, interpret, and construct line and circle graphs and show understanding of the basic concepts of probability. These students should be able to translate verbal problem situations into simple algebraic expressions and identify symbolic algebraic expressions representing linear situations.

(255) BASIC: Partial Mastery of Knowledge and Skills

The eighth-grade student performing at the basic level should be able to identify and use the correct operations for solving one- and two-step problems involving addition, subtraction, multiplication, and division of whole numbers and decimals. These students should also have an understanding of place value and order of operations, and a conceptual understanding of fractions. They should be able to use a calculator and estimation to arrive at answers to simple problems. Basic eighth-grade students can use rulers to calculate the perimeter and area of rectangular figures, and make conversions between units of measure within a given system of measurement. These students should be able to use basic geometric terms and identify elementary geometric figures. They should be able to read, interpret, and construct bar graphs and evaluate or solve simple linear equations involving whole numbers.

Grade 8 Basic: Example 1



On which day were more boxes of lemons picked than either boxes of oranges or boxes of grapefruit?

A Monday

B Tuesday

[✓ C] Wednesday

D Thursday

E Friday

F No day

G I don't know.

Grade 4: 42% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
31%	67%	79%

Grade 8: 74% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
73%	90%	97%

Grade 8 Basic: Example 2

Grade 8: 83% Correct Overall

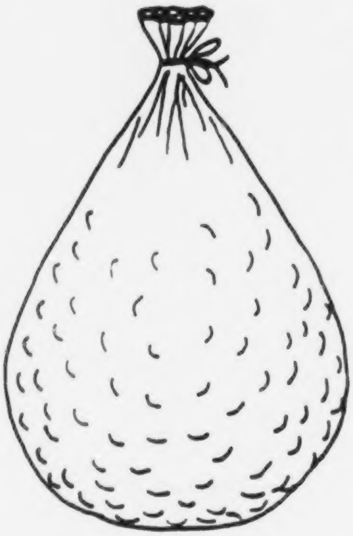
Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
84%	93%	96%

There is only one red marble in each of the bags shown below. Without looking, you are to pick a marble out of one of the bags. Which bag would give you the greatest chance of picking the red marble?

10 marbles

100 marbles



1000 marbles

[✓ A] Bag with 10 marbles

B Bag with 100 marbles

C Bag with 1000 marbles

D It makes no difference.

E I don't know.

Grade 8 Basic: Example 3

Grade 8: 77% Correct Overall

What is the value of $n + 5$ when $n = 3$?

Answer: 8

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
74%	95%	95%

Grade 8 Proficient: Example 1

Grade 8: 59% Correct Overall

In the model town that a class is building, a car 15 feet long is represented by a scale model 3 inches long. If the same scale is used, a house 35 feet high would be represented by a scale model how many inches high?

A $\frac{45}{35}$

B 3

C 5

[✓ D] 7

E $\frac{35}{3}$

Did you use the calculator on this question?

○ Yes ○ No

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
50%	84%	99%

Grade 8 Proficient: Example 2

The weight of an object on the Moon is $\frac{1}{6}$ the weight of that object on the Earth. An object that weighs 30 pounds on Earth would weigh how many pounds on the Moon?

Answer: 5

Did you use the calculator on this question?

○ Yes ○ No

Grade 8: 49% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
36%	81%	99%

Grade 8 Proficient: Example 3

If $\frac{2}{25} = \frac{n}{500}$ then $n =$

A 10

B 20

C 30

[✓ D] 40

E 50

Grade 8: 49% Correct Overall

Percent Correct At Each Achievement Level

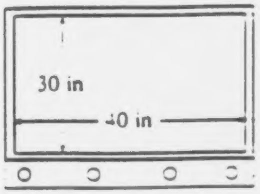
Basic	Proficient	Advanced
36%	73%	94%

Grade 12: 63% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
54%	89%	96%

Grade 8 Advanced: Example 1



What is the diagonal measurement of the TV screen shown in the figure above?

A 25 inches

B 35 inches

[✓ C] 50 inches

D 70 inches

E 1,200 inches

Grade 8: 25% Correct Overall

Percent Correct At Each Achievement Level

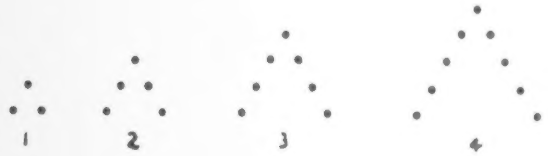
Basic	Proficient	Advanced
16%	40%	61%

Grade 12: 43% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
26%	76%	98%

The next two questions refer to the following pattern of dot-figures.



Grade 8 Advanced: Example 2

If this pattern of dot-figures is continued, how many dots will be in the 100th figure?

A 100

B 101

C 199

D 200

[✓ E] 201

Grade 8: 34% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
23%	47%	81%

Grade 12: 49% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
36%	77%	94%

Grade 8 Advanced: Example 3

Explain how you found your answer to the question above.

Answer: $(N \times 2) + 1$

Grade 8: 15% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
5%	24%	54%

Grade 12: 27% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
12%	51%	83%

Exhibit 3: Levels of Mathematics Achievement for Grade 12

(358) ADVANCED: Superior Performance

Twelfth-grade students who are performing at the advanced level should be able to investigate numerical relationships and determine the validity of conjectures involving number theory concepts such as parity (odd, even) and divisibility. These students should be able to establish procedures for the comparison and conversion of measurements of length, area, volume, and capacity. These students should understand the Pythagorean theorem and its applications, as well as use of coordinate geometry to represent relationships and solve problems. These students should also be able to graphically describe data for a situation, as well as provide numerical measures of central tendency (mean, median, and mode) and variability. Advanced twelfth-grade students should be able to apply probability and statistics concepts in reasoning about population characteristics based on information derived from a sample, including judging the adequacy of the sample. They should also be able to determine the probability of diverse events. These students should be able to translate information about linear situations from verbal or tabular forms to equations and analyze, verbally or in writing, the nature of relationships involving change in the values of the variables involved. These students should also be able to solve linear equations, inequalities, and systems of two equations in two variables, as well as evaluate a linear function and relate the value to a point on a graph of the function.

(330) PROFICIENT: Solid Academic Performance

Twelfth-grade students who are performing at the proficient level should have considerable command of the use of number and operations involving all forms of real numbers. In particular, these students should be able to represent problems involving integers, decimals, and fractions using symbols or graphs. These students should also be able to select, interpret, and use measurement relationships and formulas in problem situations. They should be able to make and evaluate conjectures about the properties of geometric figures. Proficient twelfth-grade students should be able to relate data about chance to physical models and use such models to solve problems. These students should be able to use coordinate systems on a number line to represent solutions to one-variable inequalities and use ordered pairs to describe locations in the plane.

(282) BASIC: Partial Mastery of Knowledge and Skills

Twelfth-grade students who are performing at the basic level should demonstrate conceptual and procedural understanding of whole numbers, integers, fractions, and decimals and use them when solving routine problems. They should understand and apply measurement concepts and skills, including estimation, and solve routine problems involving time, money, and length. They should also be able to read scale drawings and use formulas to find areas and volumes. Basic twelfth-grade students should be able to identify a wide range of geometric figures, describe their characteristics, and solve problems involving angle measurements and similar triangles. These students should be able to interpret data in a variety of settings, including charts, tables, and graphs. Their understanding of chance should include the ability to select favorable outcomes to a situation and find the probability of an event in a setting involving a small number of outcomes. They should also be able to simplify and evaluate simple linear expressions and solve simple one-step linear equations and inequalities.

Grade 12 Basic: Example 1

Grade 12: 79% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
76%	93%	96%

POPULATIONS OF DETROIT AND LOS ANGELES
1920 - 1970

Year	City	
	Detroit	Los Angeles
1920	950.000	500.000
1930	1.500.000	1.050.000
1940	1.800.000	1.500.000
1950	1.900.000	2.000.000
1960	1.700.000	2.500.000
1970	1.500.000	2.800.000

How many more people were living in Los Angeles in 1960 than 1940?

A 100,000

B 500,000

C 800,000

[✓ D] 1,000,000

E 2,500,000

F I don't know.

Grade 12 Basic: Example 2

If the diameter of a circle is 30 centimeters, what is the radius of the circle?

A 10 cm

[✓ B] 15 cm

C 60 cm

D 90 cm

E 180 cm

Did you use the calculator on this question?

○ Yes ○ No

Grade 12: 80% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
74%	98%	100%

Grade 12 Basic: Example 3

How many hours are equal to 150 minutes?

A $1\frac{1}{2}$

B $2\frac{1}{4}$

C $2\frac{1}{3}$

[✓ D] $2\frac{1}{2}$

E $2\frac{5}{6}$

Grade 8: 59% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
53%	76%	98%

Grade 12: 74% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
72%	87%	92%

Grade 12 Proficient: Example 1

If $f(n) = n + 5$, what is the value of $f(3)$?

Answer: 8

Grade 12: 52% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
37%	90%	98%

Grade 12 Proficient: Example 2

The perimeter of a square is 24 centimeters. What is the area of that square?

[✓ A] 36 square cm

B 48 square cm

C 96 square cm

D 576 square cm

E I don't know.

Grade 12: 45% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
20%	89%	98%

Grade 12 Proficient: Example 3

What percent of 175 is 7?

[✓ A] 4%

B 12.25%

C 25%

D 40%

Did you use the calculator on this question?

○ Yes ○ No

Grade 12: 49% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
33%	79%	93%

Grade 12 Advanced: Example 1

Grade 12: 10% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
3%	16%	45%

A contractor is building 5 different model homes on 5 adjacent lots on one side of a street. If 1 house is to be built on each lot, how many different arrangements of the 5 houses are possible?

[✓ A] 120

B 60

C 25

D 10

E 5

Did you use the calculator on this question?

○ Yes ○ No

Grade 12 Advanced: Example 2

Grade 12: 26% Correct Overall

Percent Correct At Each Achievement Level		
Basic	Proficient	Advanced
17%	36%	70%

Suppose that $a_1, a_2, a_3, \ldots$ is the sequence of numbers such that $a_1 = 3$, $a_2 = \sqrt{a_1} + 1$, $a_3 = \sqrt{a_2} + 1$, and, in general, $a_{n+1} = \sqrt{a_n} + 1$ for all $n \geq 1$. To the nearest hundredth, the value of a_5 is

A 1.63

[✓ B] 2.62

C 2.73

D 3.24

E 5.73

Did you use the calculator on this question?

○ Yes ○ No

Grade 12 Advanced: Example 3

A savings account earns 1 percent interest per month on the sum of the initial amount deposited plus any accumulated interest. If a savings account is opened with an initial deposit of $1,000 and no other deposits or withdrawals are made, what will be the amount in this account at the end of 6 months?

A $1,060.00

[✓ B] $1,061.52

C $1,072.14

D $1,600.00

E $6,000.00

Did you use the calculator on this question?

○ Yes ○ No

Grade 12: 15% Correct Overall

Percent Correct At Each Achievement Level

Basic	Proficient	Advanced
8%	21%	55%

7. Analysis of Achievement Level Ratings - Validation/Replication

7.1 Overview of Rounds One and Two Ratings

Tables 40 to 60 in appendix I contain the average achievement level ratings of judges on the first two rounds of ratings for all 21 blocks of items (7 blocks/grade level). One trend in the data is clear: The second set of ratings dropped by an average of 3 to 4 percent. This drop was due to the influence of the actual item p-values which were given to the judges prior to their completion of the second round of ratings. A second trend in the data was that the variability in the expected proportion-correct scores for BASIC, PROFICIENT, and ADVANCED students (mean item ratings) increased when judges had access to the actual item p-values.

7.2 Comparisons of Achievement Levels Across Sites--Block Level

Tables 61 to 63 contain the achievement levels for marginally BASIC, PROFICIENT, and ADVANCED students in each block of items for judges at each site for the first round of ratings. The ratings are reported at the block level rather than the booklet level to increase the sample size and to make any comparisons over sites more meaningful. The tables also contain the means and standard deviations of the block achievement levels for BASIC, PROFICIENT, and ADVANCED students after each round of ratings. In view of the modest number of items/block (about 15-20), and the small number of judges at each site, the variability in the achievement levels among sites seemed small. Also, it was clear that (generally) achievement levels dropped a few percentage points on the second round in all sites. There was more agreement in the achievement levels on the second round than the first (though there were many exceptions), especially at the ADVANCED level.

7.3 Final Round Achievement Levels

Table 64 provides a complete summary of the final achievement levels at each grade level at each site as well as the achievement levels set by the total group of judges. There was little evidence of any skewness in the distributions of judges' achievement levels (unlike the findings in phase 1). And, though the sites cannot be considered to be replications because regional differences cannot safely be assumed to be zero, in only 4 comparisons (out of 36) did a site achievement level on the final round differ by more than 5% from the average achievement level. (At grade 4 BASIC in Connecticut the difference was -6.9%; and in California, the difference was 8.4%. At grade 12 BASIC in Florida, the difference was -8.3%; and at grade 8 PROFICIENT in Michigan, the difference was 6.0%.) For five of the nine achievement levels, the maximum difference among the four sites (lowest to highest) was less than 5%. Results were the most stable at the ADVANCED level and the least stable at the BASIC level. In fact, at the BASIC level, the amount of variability across the four sites appeared substantial and troublesome. The explanation is unknown at this time. In view of the fact that the pattern appeared at all grade levels, problems with the definition of BASIC itself is a possible explanation. Another possibility is that there were real regional differences in the definition of BASIC students. Methodological problems such as the non-uniform distribution of booklets (which varied in their difficulty) across sites is another possible explanation.

Confidence level data for the judges' final ratings appear in tables 65 to 67. A 4-point rating scale was used: 1 = not confident; 2 = somewhat confident; 3 = confident; and 4 = very confident. (The rating form appears in appendix D.) The typical mean rating for an achievement level at a grade level at a site exceeded 3.0. Confidence levels were highest at grade 12. ADVANCED levels were judged more confidently than the PROFICIENT levels which in turn were judged more confidently than the BASIC levels.

7.4 Evaluation of the Achievement Level-Setting Process

Tables 66 and 67 contain the results of the survey of the judges about their perceptions of the process. (A copy of the survey appears in appendix D.) Highlights of the evaluation follow:

1. Seventy-six percent judged the training to be appropriate; 23% judged it to be somewhat appropriate.
2. Sixty percent said they were clear about the definition of BASIC; 35% said they were somewhat clear. At the PROFICIENT level the ratings were considerably better with 74% clear and only 25% indicating somewhat clear. At the ADVANCED level, the results were considered better again with 81% clear, and only 19% somewhat clear.
3. In terms of the time allotted to complete the work, 83% felt the timing was right; 11% felt not enough time was allotted.
4. Ninety-eight percent of the judges indicated that their level of understanding of the process was medium or high.
5. Primary factors in the judges' ratings were (1) the definitions (89%), (2) item content (83%), (3) perceptions of item difficulty (92%), and (4) actual item performance (74%). About half the judges indicated their final ratings were influenced by other judges at their grade level. Judges from other grade levels did not appear to be a factor in the achievement levels.
6. On the question of usefulness of the resulting achievement levels, 87% felt "Definitely Yes" (36%) or "Probably Yes" (51%); 11 percent were unsure.

Table 67 provides the statistics on the demographic makeup of the judges. Perhaps the important points to highlight are the very high percentage of educators/math educators (87%) and the diversity of the environments, grade levels, and types of students they teach.

7.5 Evaluation of the Expanded Definitions

During the achievement level-setting process, a supplemented subgroup of the Vermont panel reviewed the item ratings and prepared descriptions of BASIC, PROFICIENT, and ADVANCED content. These definitions included mathematical skills and behaviors that would be mastered by students at each level. The judges were asked to indicate whether or not they thought particular skills should be included in the definitions. A summary of the judges' responses at grades 4, 8, and 12 is contained in tables 68, 69, and 70, respectively. With only minor doubts or exceptions, the judges approved the list of skills. They were, with very few exceptions, unable to suggest the addition of new skills to the lists.

7.6 Additional Analyses

The Technical Review Panel criticized the third round of ratings in the Vermont/Washington study because of the high correlations between the actual item p-values and the expected item p-values as set by the judges. Tables 71, 72, and 73 provide a complete set of correlations (at the block level) of the first and second rounds of ratings and the actual p-values for grades 4, 8, and 12, respectively. In all instances (63), the correlations reflect the substantial influence of the actual item p-values on the ratings. On the other hand, correlations between the first and second sets of ratings were very high too. At grade 4, the lowest correlation (of 21) was 0.75. At grade 8, the lowest correlation (of 21) was 0.71, and the second lowest correlation was 0.86. At grade 12, the lowest correlation was 0.95. The correlation between the first round of ratings (which was completed without knowledge of the actual item p-values) and the actual item p-values ranged from 0.44 to 0.89 at grade 4, 0.50 to 0.91 at grade 8, and 0.76 to 0.93 at grade 12. Clearly, the high pattern of correlations observed between the round two ratings and the actual item p-values was not due solely to the presence of the item p-values in the process. Judges seemed capable (even at round one) of judging item difficulty and incorporating it into their process of item ratings.

Table 74 provides the results of a second analysis: a comparison of achievement levels of educators and noneducators. This analysis was inspired by analyses of the phase 1 data, where an attempt was made (but later rejected) to correct the grade 4 achievement levels because of the lack of noneducators at the second meeting. Unfortunately, as is clear from Table 74, the number of noneducators in the phase 2 study was too small to conduct a stable comparative analysis. However, one can notice a trend in the results for noneducators to set slightly higher achievement levels (but inconsistencies in this trend were apparent too).

One of the main problems with the Phase 1 activities was that the grade 8 results seemed to be inconsistent at round three with the results at other grade levels. The problem was identified by analyzing judges' ratings on the items common to two or three grades (see tables 10 to 14 in appendix F).

In fact, one of the primary reasons for reconvening the second meeting in Washington was to address this problem of incoherence in the results across grade levels. Tables 75, 76, and 77 provide the actual item p-values and, more importantly, the expected item p-values for BASIC, PROFICIENT, and ADVANCED levels on the common items. Of the 294 possible between grade comparisons (120 in table 75, 60 in table 76, and 114 in table 77), only one reversal was found, and the achievement levels on the common items showed substantial increases across grade levels. The evidence seemed clear that, using the common items only, there was coherence in the achievement levels across grade levels. The weakest distinction appeared to be between grades 8 and 12 ADVANCED, though this finding would not necessarily generalize to the larger pools of test items at grades 8 and 12, when reporting achievement levels on the NAEP reporting scale. Actually, the distinctions observed among the other achievement levels and grade levels would not necessarily generalize either. Perhaps the main point of this analysis is that, on the basis of the data in tables 75 to 77, there is evidence for the coherence of achievement levels over grades 4, 8, and 12.

7.7 Comparison of Phase 1 and 2 Final Achievement Levels

The final recommended achievement levels from the two phases of the process were as follows:

Grade	Level	Vermont/Washington (N=38)	Replication/Validation (N=211)
4	Basic	51%	45%
	Proficient	76	68
	Advanced	91	87
8	Basic	60	48
	Proficient	80	72
	Advanced	92	89
12	Basic	53	47
	Proficient	79	73
	Advanced	90	88

In all cases, the achievement levels were lower in the replication/validation phase than in the Vermont/Washington phase. An analysis of the decreases revealed that the average decrease over the nine achievement levels was 6%, largest at grade 8 BASIC (12%), larger in general at the BASIC (8%) and PROFICIENT (7%) levels than at the ADVANCED level (3%) and larger at grade 8 (8%) than at grade 4 (6%) or grade 12 (5%).

The most plausible explanation for the decrease was the change in the demographic characteristics of the judges who set the achievement levels. The replication/validation phase consisted of mainly classroom teachers, whereas the Vermont/Washington phase included more mathematics supervisors, coordinators, university professors, school administrators, and more noneducators. There were also changes in the process and in the environment, which could have been influential on the ratings process. For example, at the Vermont/Washington meetings, the environment was "electric" with government officials, ETS staff, evaluators, a newsperson, and other dignitaries being present. A more calm atmosphere prevailed at the replication/validation meetings.

8. Additional Topics

8.1 Introduction

The purpose of this chapter is to discuss the important issues that were raised during the standard-setting process by outside consultants and or stakeholder groups. The issues fall into three categories: (1) discrepancy between time of testing and time of standards; (2) corrections for guessing; and (3) estimating variability.

8.2 Discrepancy in Time of Testing vs. End-of-Year Standards

The judges in Vermont raised the issue of time of testing versus time of standards. Essentially, the arguments are as follows. In setting standards, the training materials asked judges to think about the performance of examinees as they complete the grades in which they are assessed--namely, 4, 8, and 12. It simply did not make sense to attempt to have judges think about examinees in February or March of the school year (which is when the assessment is given). But rather, as students exit fourth grade, or eighth, what should they be expected to be able to do? This makes the task for the judges clearer, but it creates the discrepancy problem, since the assessment was administered in the winter of 1990 (between February 5 and March 2, 1990).

The most obvious and relatively straightforward resolution of this problem is to simply make a statistical adjustment of the end-of-year cut scores to accommodate a winter performance estimate. Beginning in 1990, the three age/grade samples were based on calendar-year definitions of age (and consequently modal grade was adjusted) in order to ensure that there was 4 years of growth between the three age/grade samples. Assuming a linear relationship between fourth and eight grade performance, for example, the cut scores could be adjusted down by one-twelfth of the difference of the means (4 months difference out of 48 months). This would make a very slight change in the cut scores of about 2 to 3 scale score units, causing a slight adjustment in the percentage of students judged to be BASIC, PROFICIENT, or ADVANCED. In 1990 this adjustment was not made.

8.3 Correction for Guessing

The guessing factor is an issue that was raised by the Vermont panel, and by the Technical Review Panel evaluation. It is not a concern that the architects of the process gave advanced thought to simply because it is seldom attended to in the standard-setting process. Taking guessing into account is a far more critical issue when the cut scores are at or near "chance level" scores, and this is almost never the case. However, in setting achievement levels, the BASIC cut scores could be approaching the low end of the NAEP scale, and consequently, guessing becomes an important factor.

There are a number of ways to approach a solution. One method would be to include a consideration of guessing in the training of the judges, and have each judge take this into account as they make their judgments on each item or on the item pool as a whole. A second method is to make a statistical adjustment in the judges' ratings, much the same way a "guessing formula" is applied to the scoring of tests.

It is the judgment of the authors that neither of these approaches is an acceptable solution. In the first case, it is not clear that training judges to consider guessing will result in a standardized approach to the problem. Different judges will interpret the training differently, and make corrections of differing magnitudes, perhaps apply them unequally to different items, and almost certainly apply them for different reasons. The resulting levels from each judge would be uninterpretable and mathematically intractable.

In the second case, it is not clear what statistical adjustment should be made, and of what magnitude. Would the adjustment be the same for all judges? Would the adjustment apply equally to all items? With as many unknowns, it simply seems better to suggest that, until some future research studies examine these questions, no adjustments should be made. This was the position that advisers to the project took this year.

It should be noted that NAEP currently employs an Item Response Theory model in which guessing is one of the item parameters and is taken into account in the estimation of proficiencies and the development of the scale.

8.4 Estimating Variability

There are several sources of error that can contribute to instability in the achievement levels. Interjudge and intrajudge inconsistency are primary sources of error as well as fluctuations due to sampling and the composition of the panel of judges.

Due to constraints of time and resources, it was not possible to examine fully each of these error sources in the 1990 process. Interjudge consistency was examined in terms of measures of central tendency and variability within the distribution. Intrajudge consistency was examined by an analysis of judges' ratings on common items, and correlations of estimated probabilities with item p-values.

The 1992 process will attempt to look at these and other sources of errors as well. The 1992 design will give particular attention to interrater reliability and intrajudge consistency, and will identify and analyze other potential sources of error that could contribute to instability in the achievement levels.

9. Conclusions and Recommendations

9.1 Summary

Setting achievement levels on the National Assessment of Educational Progress has been a landmark effort. Never before has there been an initiative of this magnitude, involving a national survey. However, precisely because of its magnitude and implications, any future efforts for setting achievement levels on NAEP must be more trouble-free than either phases 1 or 2 of this process. This section of the report, therefore, will summarize what the authors believe to be the primary advantages and disadvantages of both phases 1 and 2 so that the advantages can be incorporated and improved upon in any future achievement levels-setting efforts, and the disadvantages minimized, if not eliminated.

9.2 Advantages and Disadvantages of Phase 1

One of the most notable advantages of Phase 1 was the diversity of the panel of judges who participated in the process. The sample of judges, drawn from candidates provided by major national organizations, were, in many cases, national figures in their own right. Their talent and expertise provided a broad, comprehensive view of mathematics education. The panel also included full participation by the noneducator segment, deemed very important by Board policy.

Another distinct advantage was the review of the entire item pool by all panel members. With the exception of the Higher Order Thinking Skills and Estimation items (which should not have been included, perhaps), judges reviewed each item in the context of all other items. This allowed the judges to have a complete picture of what was being asked of examinees in responding to the assessment.

In addition, the training materials were carefully prepared and reviewed by numerous individuals qualified to make suggestions for improvement. The briefing materials covered a broad range of topics to bring panelists "up to speed" as quickly as possible.

What were the disadvantages of phase 1? There were numerous problems. First, judges were not comfortable with the generic definitions provided by the Board for conceptualizing the three levels (BASIC, PROFICIENT, and ADVANCED). The definitions were not sufficiently operationalized to allow the judges to have a common understanding of what the Board meant by the levels. This caused problems in the rating tasks, especially at the BASIC levels, which only showed up after the first rounds of data were collected.

Second, as mentioned above, many of the judges were candidates suggested by major, national organizations, and, in many instances, they were representing their constituencies and wanted to do that well. In some ways, the task of rating test items was too mundane. The judges were extraordinarily committed, but because of their professional stature tended to be more outspoken about the quality of the item pool and other aspects of the process that were considered "givens." Consequently, much time was lost responding to questions and comments that were not germane to achieving the goals of the meeting.

Related to this issue was another problem that was not anticipated, namely, confusion between item difficulty and item appropriateness. Not all judges were happy with the quality of the item pool, and thus, claimed many items were inappropriate: either they were inappropriately worded, out of sequence for the grade level, or otherwise faulty. If a judge thinks an item is easy but inappropriate, how should he or she rate the item? The distinction between item difficulty and item appropriateness was not resolved until the second meeting of the group in Washington, DC.

Because of the tight timeliness, pilot testing the training materials and timing of tasks was not done. This caused several problems that should have been anticipated. Two days was not sufficient time to complete all the tasks required in the process; generic definitions were not sufficiently operationalized; there was confusion over the rating task itself; and there was variability in the rating process among some of the 12 subgroups of judges caused by a lack of attention to the instructions of the training staff.

Finally, in an effort to be open and "above-board" about the process, the number of observers of the process was not restricted, as long as the observer was willing to sign a non-disclosure form ensuring item security. This, in effect, resulted in a "fishbowl" atmosphere. Judges and staff felt as if they were "being watched," and this environment gave some of the more vocal participants a platform for airing their views and opinions.

9.3 Advantages and Disadvantages of Phase 2

The advantages of phase 2 were in large measure due to improvements in the process made as a result of the phase 1 experience. First, the judges accepted the Board's generic definitions, and felt comfortable applying them in the achievement level-setting process. Second, the training via videotape was more than adequate; judges understood their task; directions were clear; and the materials were reasonable to move through. The training tape allowed replicability across sites since the essential part of the training was standardized in content and presentation. The "fish bowl" atmosphere was also gone for the most part which helped considerably.

Judges were asked to rate only about 40% of the total item pool, i.e., a student booklet consisting of 3 blocks (with about 20 items per block). The sample of judges was increased to account for the reduced number of items per judge, so that each item received approximately the same number of ratings as in phase 1 (about 25). This reduction in the size of the item pool allowed the rating process to be completed in one day at each site. The composition of the judges changed as well. The four participating states were quite helpful in selecting teachers and noneducators who met the requirements set out in the Board's replication/validation plan. This brought with it the distinct advantage of a group of raters who did not have national agendas; they were based in classrooms, had an understanding of what students' capabilities are, and were more focused on the task.

Were there disadvantages in phase 2? No process is without its problems. And this one is no exception. First, because the item pool was matrix sampled, 7 booklets per grade, and 21 booklets per session, in a group of 21 judges per grade level (and many groups were not this large), only 3 judges rated the exact same item set. Therefore, discussion about the items and how they were rated was difficult. This was an important loss, since discussion many times reduces variability in the ratings due to random, careless errors, confusion, or misunderstandings on the part of the judges.

Second, participants came in "cold" to the process for only one day. There was no time to bring judges "up to speed" on NAEP, the Board, and barely, even the achievement level setting process itself. Briefing materials may have helped if we had been able to get them to participants in sufficient time. However, the selection process by the states was occurring during spring break, the Easter holidays, and the American Educational Research Association/National Council on Measurement in Education convention. Coordination was limited because of extenuating circumstances beyond the control of the project.

9.4 Recommendations for Future Efforts

The recommendations are clustered into three broad categories: the sample of judges, the item rating tasks, and data analysis.

Sample of Judges. The number of judges needed is a function of how much of the item pool is rated by each judge. It is recommended that the number of raters be such that each item in the pool has approximately 25 ratings. Based on the experience of phases 1 and 2, this seems to be a reasonable number that will yield fairly stable estimates of the achievement levels, and less than which runs the risk of adding significant sampling error to the overall standard errors.

The background of the judges (educator and noneducator) has been determined by Board policy. However, it is clear from the experience of phase 2 that classroom teachers are probably in the best position to judge whether or not an examinee should get an item correct, and thus, meet the definitions of BASIC, PROFICIENT, or ADVANCED. We would recommend that the 70% educator segment be highly concentrated (as high as 50%) with judges having similar characteristics to the replication/validation sample. It is also recommended that differentiated briefing and training materials be used with the two segments of the panel, i.e., educator and noneducator. The noneducator segment particularly needs to be acquainted with NAEP, Board policy, large-scale assessment, and other relevant topics. This could be achieved by having a pre-session training (one day) for the noneducator group.

Item Rating Tasks. It is fairly clear that the generic definitions provided by Board policy to guide the achievement level-setting process are insufficiently developed for judges to use in rating items. It is recommended that these definitions be operationalized within the specific framework of each content area. Criterion-referenced statements based on the frameworks and test specifications documents which elaborate BASIC, PROFICIENT, and ADVANCED for each grade level are essential. This is probably best achieved by the panel of judges participating in the standard-setting process, prior to the item-rating task.

The cumbersomeness of the item rating task is related to the number of items per judge. Rating the full item pool has the advantage of sponsoring good discussion to reduce variability. It has the disadvantage of being overwhelming when the number of items exceeds reasonable limits (150-200 items). The 1992 mathematics assessment may present just such a problem, since the number of items in the pool has increased substantially--from 7 blocks per grade to 13 per grade. It is recommended, therefore, that an item-sampling procedure be devised that would allow maximum discussion while minimizing the burden. It is also recommended that the number of rounds of data collected be consolidated to take advantage of the amount of time required to complete the process.

The training of judges was assisted by the videotape approach used in phase 2. If in future efforts several different sites will be used, it is highly recommended that consideration be given to standardizing the training presentations through video or some other procedures that ensure consistency and clarity from site to site. This is particularly important if the data collected at various sites will be aggregated to composite results. It is also an efficient method for conducting the common training across various subject areas.

It is highly recommended that all procedures be piloted before implementation. If this process had enjoyed the luxury of pilot testing before phase 1 many of the problems encountered could have been anticipated and corrected. Finally, the authors recommend that reasonable restraint be used in inviting observers to the standard-setting meetings. The "open-door" policy of 1990 did not contribute positively to the atmosphere of the meetings, or to accomplishing the goals of the meetings in a timely fashion.

Data Analysis. The data analyses completed during phases 1 and 2 was extensive. The analyses examined group ratings, subgroup ratings, common-item analysis, measures of variability, demographic subgroup differences, differences between and among rounds of data, and content/process breakdowns. It was on the basis of such analyses that technical decisions were made throughout the process, and on which the Board based its final decision regarding the levels. Certainly, future efforts should include similar analyses. It is also recommended that additional analyses should be devoted to identifying and minimizing alternate sources or error, and to additional ways of reporting the achievement levels appropriately.

Finally, matters such as the need for achievement levels on the content/process scales must be considered as well as the implications of multidimensionality in the item pool on IRT scaling and achievement level setting.

10. References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), Educational Measurement. Washington, DC: American Council on Education.

Busch, J. C., & Jaeger, R. M. (1990). Influence of type of judge, normative information, and discussion on standards recommended for the National Teacher Examinations. Journal of Educational Measurement, 27(2), 145-163.

Fitzpatrick, A. R. (1989). Social influences in standard setting: The effects of social interaction on group judgments. Review of Educational Research, 59(3), 315-328.

Forsyth, R. A. (1991). Do NAEP scales yield valid criterion referenced interpretations. Educational measurement: Issues and practice, 10, 3-9, 16.

Hambleton, R. K., & Powell, S. (1983). A framework for viewing the process of standard-setting. Evaluation and the Health Professions, 6(1), 3-24.

Johnson, E. G. (1991, April). Defining levels on the 1991 mathematics composite. Paper presented at the meeting of the American Educational Research Association, Chicago.

Appendix A

Panelists in Vermont/Washington, DC

Appendix A

Panelists in Vermont/Washington, DC[1]

Judy Adams	Laramie Public Schools, Laramie, WY
Peter Andre	U. S. Naval Academy, Annapolis, MD
Linda Barnett	Council for Talented Youth, Baltimore, MD
Bruce C. Burt	East Bradford Elementary School, West Chester, PA
M. Blouke Carus	Carus Corporation, Peoria, IL
Nancy Cetorelli	Assistant Superintendent, Huntington, CT
Donald Chambers	Wisconsin Department of Education, Madison, WI
Gordon Clem	Choir of St. Thomas School, New York, NY
Nora Cronin, PBVM	Loyola High School West, Wichita, KS
F. Joe Crosswhite	Flagstaff Public Schools, Flagstaff, AZ
Jodi Crowe	Beaverton Public Schools, Beaverton, OR
Rubye S. Dobbins	Arlington School Board, Arlington, TN
John Dossey	Illinois State University, Normal, IL
Carl Downing	Central State University, Edmond, OK
Paula B. Duckett	River Terrace School, Washington,DC
Linda Durant	South Carolina Educational Television, Columbia, SC
Robert Gabrys	Maryland Department of Education, Baltimore, MD
Mardi Gale	California Public Schools, Beverly Hills, CA
Arthur Griffith	Legal Services, Charlotte, NC
Terrance Henry	U.S. Military, Chicago, IL

---

[1] The number of participants listed and the number appearing in the data tables (appendix F) are discrepant by one participant who did not wish to have his/her name appear in this listing.

Richard L. Hinman	Pfizer Chemical Co., Groton, CT
Susan Hooker	Motorola Corporation, Schaumburg, IL
Margaret Ingram	Beachland Elementary School, Vero Beach, FL
Mary Jane Raeihle, SSJ	St. John Baptist School, Brooklyn, NY
Margaret Kaduce	Chippewa Falls Middle School, Chippewa Falls, WI
Ann P. Kahn	National Academy of Sciences, Washington, DC
James W. Keefe	National Association of Secondary School Principals, Reston, VA
Robert Dale Keefer	Wichita High School West, Wichita, KS
John Kenelly	Clemson University, Clemson, SC
Robert Kenney	Vermont Department of Education, Montpelier, VT
Jeanne P. Klein	Council for American Private Education, Apple Valley, MN
Mary Harley Kruter	National Academy of Science, Washington, DC
Karen R. Kundin	Kachina School, Glendale, AZ
Zoe Leimgruebler	Oklahoma Department of Education, Oklahoma City, OK
Sharon Johnson Lewis	Detroit Public Schools, Detroit, MI
Steve Lienwand	Connecticut Department of Education, Hartford, CT
Mary Lindquist	Columbus College, Atlanta, GA
Harvey Long	I.B.M., Rockville, MD
Delores McGhee	Atlanta School Board, Atlanta, GA
Laurietta McNealy	Mays Middle School, Miami, FL
Gloria Moretti	San Matel Public Schools, San Matel, CA
James B. Olsen	WICAT Systems, Orem, UT
Arnold Packer	Department of Labor, Washington, DC

Steffan Palko	Timbers Oil Company, Fort Worth, TX
Tej Pandey	California Assessment Program, Sacramento, CA
Carole Perlman	Chicago Public Schools, Chicago, IL
Yolanda Rodriguez	Cambridge Public Schools, Cambridge, MA
Thomas A. Romberg	Wisconsin Center for Educational Resources, Madison, WI
Edward Schwarze	Caterpillar, Inc., Peoria, IL
Nannette Seago	Mission Middle School, Riverside, CA
Joan Sextro	New Trier High School, Winnetka, IL
Dorothy Strong	Chicago Public Schools, Chicago, IL
Marylin N. Suydam	ERIC Clearinghouse, Columbus, OH
Judith Thayer	New Hampshire State Board of Education, Manchester, NH
Susan Thomas	San Antonio Public Schools, San Antonio,TX
Judith Trowell	Little Rock Public Schools, Little Rock, AR
Harry J. Vriend	West Side Christian School, Grandville, MI
John B. Walsh	Boeing Aerospace, Seattle, WA
Charles Watson	Arkansas Department of Education, Little Rock, AR
Vernon Williams	H.W. Longfellow Intermediate School, Falls Church, VA
Mary Jackson Willis	Governor's Office, Columbia, SC
Robert Ziomek	Cedar Rapids Public Schools, Cedar Rapids, IA

Appendix B

Training Manual for Phase I

Setting Achievement Levels

for the

1990 NAEP Mathematics Assessment

Handbook for Judges

August 1990

Table of Contents

Background and Rationale 110

Introduction 111

Agenda 113

The Standard-Setting Method 113

Working Description of the Basic, Proficient, and Advanced Student 117

Practice Exercise 118

Standard-Setting Procedure 118

Acknowledgment 121

Appendix A: Practice Item Rating Form 122

SETTING ACHIEVEMENT LEVELS

BACKGROUND AND RATIONALE

Among the most significant responsibilities of the National Assessment Governing Board are (1) taking appropriate actions . . . to improve the form and use of the National Assessment; and (2) setting "appropriate achievement goals" for each grade and subject tested under the National Assessment of Educational Progress (NAEP). The two responsibilities fit well together. By defining levels of appropriate achievement on the National Assessment, the Board will increase greatly the significance and usefulness of NAEP results to educators, policymakers, and the American public.

The statute (P.L. 100-297) creating the Board assigns to it certain explicit responsibilities:

- Taking appropriate actions needed to improve the form and use of the National Assessment;
- Developing . . . standards for analysis plans and for reporting and disseminating (NAEP) results;
- Developing standards and procedures for interstate, regional, and national comparisons;
- Identifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;
- Developing assessment objectives (and) specifications;
- Devising goal statements for each learning area assessment through a national consensus approach that provides for the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the general public.

The National Assessment Governing Board is not authorized to establish any overarching national goals for education. It does have authority to define levels of achievement that will serve as "appropriate achievement goals" on National Assessment. With such achievement levels defined, NAEP results will be reported in terms that better denote the quality or value of student achievement than do the numerical scores that represent the range of student performance.

By law, the National Assessment is a survey - not a mass individual testing program - in which representative samples of students are asked questions in different academic subjects. The assessment provides information on aggregate or group performance; it is forbidden by law to report data on individuals.

Hence, the achievement levels defined by the Board will be used for reporting group data and making it more meaningful. The assessment will not become a device for certifying or classifying individual students.

In a letter to the Governing Board, Education Secretary Lauro F. Cavazos said that, by "setting achievement standards for the National Assessment," the Board "would fulfill (its) statutory responsibility . . . (under) the Hawkins-Stafford Amendments of 1988 . . . The result would be a clear definition of what constitutes grade level performance in each subject so that future National Assessment of Educational Progress (NAEP) reports could provide data on the proportion of students who achieve that standard and in what ways American students exceed or fall short."

INTRODUCTION

On August 16 and 17, you, along with 70 other educators, business leaders, and representatives of the public will be setting achievement levels or standards on the 1990 NAEP mathematics assessment for grades 4, 8, and 12. Final achievement levels or standards will be set by the National Assessment Governing Board based on your recommendations and those of the other members of the panel. The standards[1] will be used to determine the numbers of students in the nation who are meeting three levels of mathematics achievement: Basic, Proficient, and Advanced. These levels will be described in a subsequent section of this Handbook.

The task of setting standards or achievement levels on the 1990 NAEP mathematics assessment involves judgment. In fact, you and other judges at the two-day meeting have been selected to provide your best judgments to help in setting standards of performance. In the following sections of this document you will find:

- An agenda for the two-day meeting
- A description of the details of the standard-setting method
- Working descriptions of basic, proficient, and advanced students
- A practice standard-setting exercise
- An outline of the actual standard-setting procedure

Other materials will be distributed at the meeting:

1. A practice exercise which includes 10 eighth-grade test items;
2. A copy of the 1990 NAEP mathematics items with which you will work;
3. A rating form; and
4. Item statistics in the 1990 NAEP mathematics assessment.

And, with this Handbook, you have received a number of other documents in advance of the meeting to help you prepare for the two-day meeting.

---

[1] The term "standard" is found widely in the education literature and so it will be used interchangeably in this Handbook for Judges with the term "achievement levels" preferred by NAGB.

AGENDA

Wednesday, August 15

6:30 - 7:30 p.m.	Reception, Dinner Meeting, Meet the Staff, Charge from the Board

Thursday, August 16

8:30 - 9:00 a.m.	Continental Breakfast
9:00 - 10:15 a.m.	Introduction to the Process of Standard Setting Review Content Descriptions of the Assessment and the Levels of Student Achievement
10:15 - 10:30 a.m.	Break
10:30 - 10:45 a.m.	Item Security and Security Sign-Off Form
10:45 - 11:45 a.m.	Practice Standard-Setting Exercise and Discussion
11:45 - 12:45 p.m.	Independent Item Ratings (First)
12:45 - 1:30 p.m.	LUNCH
1:30 - 4:30 p.m.	Independent Item Ratings (First)

Friday, August 17

8:30 - 9:00 a.m.	Continental Breakfast
9:00 - 10:15 a.m.	Introduction of Empirical Data Independent Item Ratings (Second)
10:15 - 10:30 a.m.	Break
10:30 - 12:00 noon	Group Discussion of Item Ratings and Preparation of Final Item Ratings
12:00 - 1:00 p.m.	Check-out and LUNCH
1:00 - 3:00 p.m.	Group Discussion of Item Ratings and Preparation of Final Item Ratings
3:00 - 3:30 p.m.	Presentation and Discussion of Achievement Levels within Each Grade Level Group
3:30 - 4:00 p.m.	Presentation of Grade-Level Results, Wrap-up and Future Steps

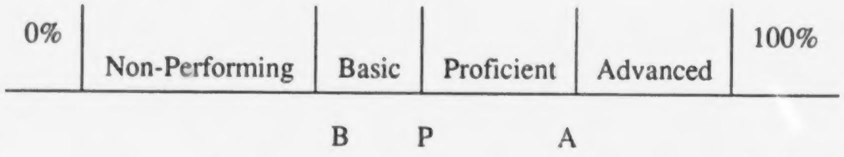
THE STANDARD-SETTING METHOD

The National Assessment Governing Board, in consultation with Ronald K. Hambleton from the University of Massachusetts and several other experts in the standard-setting field, have chosen to use a modification of the Angoff Method for setting standards (i.e., achievement levels) on the 1990 NAEP mathematics assessment. Dr. William Angoff, who introduced the method in the early 1970s, is a distinguished research scientist at Educational

Testing Service (ETS) in Princeton, New Jersey. His method is the most popular judgmental method in use today and is used by many state departments of education, credentialling agencies, and school districts.

The standard-setting method is designed to establish standards on the percent score scale to split students into four groups: non-performing, basic, proficient, and advanced. See the diagram below:

- Percent Score Scale -



Your task is to help in setting the standards or achievement levels, B, P, and A, to be used in classifying students. In finding the points B, P, and A, you must specify what you believe should be the performance of the marginally basic, marginally proficient, and marginally advanced student. Specifically, your task is to state how well these marginal students should be expected to perform on each item in the assessment. What should these marginal students know and be able to do? Remember, too, that all items in the assessment can be referenced to the objectives which appear in the Mathematics Objectives: 1990 Assessment. How this is done will be explained at the meeting.

For the Marginally Basic student, your task is to specify the probability that this marginal student should answer each item in the assessment correctly. This chance or probability for each test item can range from zero (where you would be specifying that the marginal student should have no chance of giving a correct answer) to 1.00 (where you would be specifying that the marginal student should, without a doubt, answer the item correctly).

After specifying the performance level for the marginally basic student on an item, you must provide estimates on the same item for the marginally proficient and marginally advanced student. You must have in mind a description of these three types of marginal students before you begin your ratings. These descriptions are presented in the next section. For example, you may feel that the probability should be .60 for the marginally basic student, .85 for the marginally proficient student, and .95 for the marginally advanced student.

Sometimes judges find it easier to imagine groups of 100 marginally basic, 100 marginally proficient, and 100 marginally advanced students and then specify the proportion of students in each group who should answer each item correctly. It is your choice: (1) You may specify the probability with which the minimally capable Basic, Proficient, and Advanced student should answer each item correctly, or (2) you may specify the number of students in each group of 100 who should answer the item correctly. Both ways of thinking about the rating task are acceptable. For example, saying that a single student should have a zero probability of answering an item is the same as saying that none of 100 students at the same ability level as the single student should answer it correctly. Saying that the single student has a .50 probability is the same as saying that 50 such students out of a group of 100 should answer the item correctly, and so forth. Remember, too, that your task involves stating what you believe should happen not what will actually happen.

Your standard or achievement level will be found by summing the probabilities you assign for each group to the items in the assessment and then dividing this sum by the number of items. In statistical jargon, the sum of the estimated probabilities in (say) the Proficient group should equal the expected total test score for the minimally capable performers in the Proficient group. For example, suppose you assigned ratings for the marginally proficient student of .50, .80, .80, and .90 to the items on a 4-item test. The sum is 3.00, which leads to a standard of 75% (3.0/4) on the 4-item test. Since 3.0 was the expected score on the assessment for the marginally Proficient student, it becomes the standard or cut-off score. Because each judge will produce a somewhat different standard, the standards of judges will be averaged to arrive at a final standard.

You will provide three sets of ratings. The purpose of your first set of item ratings is to determine achievement levels for students at the lowest levels (i.e., marginal) of three ability categories, Basic, Proficient, and Advanced, independent of (1) any information about how students actually performed on the mathematics assessment, or (2) the opinions of other judges. We are interested initially in your independent opinions about what you think students should know and be able to do.

Next, on Friday morning, you will be provided with some statistical information about how well students actually performed on the test items and then you will be asked to review your ratings in light of the statistical information. You may revise your first set of ratings if you feel that the achievement levels you set are too high or too low. It is not necessary for you to revise any of your ratings. Details on the item statistics will be provided at the meeting. Some practice in using the statistics will also be given. The purpose of the second set of ratings is to determine your views about what the achievement levels should be, knowing something about the current performance levels.

In the third and final stage of the item rating process, we want you to discuss your item ratings with other group members. Sometimes judges will miss an important aspect of the item or be unusually strict, or unusually lenient. Sometimes the attractiveness of a near correct answer choice is overlooked. The goal of this phase of the process is to share views about the item, the content it measures and its item statistics, and the importance of the item at the grade level where it is placed. Then judges will provide a third and final set of item ratings. The goal at this stage is not to reach consensus. It is your choice about whether or not to revise your ratings. Your final ratings will not be known to or discussed by your work group or any other members of the panel.

WORKING DESCRIPTIONS OF THE BASIC, PROFICIENT AND ADVANCED STUDENT

In applying the standard-setting method, descriptions of Basic, Proficient, and Advanced-level students are needed. These descriptions, based on discussions with mathematics educators, have been developed by the NAGB and are provided below. These descriptions will be considered in more detail when the groups begin their work. To facilitate the standard-setting process, judges at each grade level have been divided into four groups. Each group is intended to reflect the diversity of judges represented in the total group of judges.

Basic: This level denotes partial mastery of knowledge and skills that are fundamental for proficient work at each grade -- 4, 8, and 12. For 12th grade, this will be higher than minimum competency skills (which normally are taught in elementary and junior high schools) and will cover significant elements of standard high school-level work.

Proficient: This central level represents solid academic performance for each grade tested -- 4, 8, and 12. It will reflect a consensus that students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. At grade 12, the proficient level will encompass a body of subject-matter knowledge and analytical skills and cultural literacy and insight that all high school graduates should possess for democratic citizenship, responsible adulthood, and productive work.

Advanced: This highest level signifies superior performance beyond proficient grade-level mastery at grades 4, 8, and 12. For 12th grade, the advanced level will show readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, this standard may be based in part on international comparisons of academic achievement and may also be related to Advanced Placement and other college placement exams.

PRACTICE EXERCISE

During the morning of the first day, a small practice exercise will be completed using 10 grade 8 test items. You will be asked to do two things: (1) re-read the descriptions of the Basic, Proficient, and Advanced students and then (2) provide your best judgments of the performance of the three types of students on the 10 items. You will be asked to place your ratings on the Practice Item Rating Form that appears in APPENDIX A. You will use the "first rating" column on the form. The only goals of this exercise are to give you some practice in completing the rating form and in working with the three descriptions. These activities will set the stage for your work in subsequent parts of the meeting.

STANDARD-SETTING PROCEDURE

Each judge has been assigned to review test items at one of three levels: grade 4, 8, or 12. Judges have been further divided into one of four groups (of five or six participants each) at each grade level. This organization yields (approximately) 70 judges divided into 12 groups across all grade levels.

The following steps will be completed in setting standards:

1. Introduction. The 12 groups will meet and introduce themselves, then discuss and clarify the descriptions of the Basic, Proficient, and Advanced student. A moderator from each group has been identified.
2. First Set of Ratings. With a copy of the assessment and rating form in hand, each judge will provide his/her first set of item ratings. Discussion among group members may take place in order to clarify points about the rating task, but otherwise discussion should be kept to a minimum. To the extent possible, your first set of ratings should be totally independent of other judges.
3. Second Set of Ratings. The second set of item ratings will also be made independent of other judges, but this time judges will be provided with item statistics information based on an administration of the test items to a nationally representative sample of students in the spring of 1990. These item statistics will basically inform judges about current student performance.
4. Discussion of Ratings. A discussion of your first and second set of ratings will take place in each group, moderated by a member of the group. The discussion will center on your first and second sets of ratings. The moderator's task is to coordinate the discussion. For each item, high and low ratings for each type of student will be identified and reasons discussed for these ratings, along with other pertinent points about the item. Following the discussion, judges will provide a third set of ratings. Then, discussion will shift to the next item and so on until all items have been rated a third time. After the last item has been reviewed, the standard for the Marginally Basic, Proficient, and Advanced student will be calculated.

5. Completion of Rating Form. The item rating form should be returned, along with the NAEP mathematics assessment booklet, to the moderator.

6. Grade-level Meeting. A meeting will be convened of the four groups at each grade level and the basic, proficient, and advanced achievement levels for each working group will be presented and discussed. This meeting will be convened by a member of the staff. Recommended achievement levels will be considered and discussed.

7. Total Group Meeting. The total group of judges will be reconvened for the purpose of presenting and discussing the recommended achievement levels at grades 4, 8, and 12. Wrap-up and future steps will also be discussed.

Acknowledgment

The National Assessment Governing Board is grateful to the following individuals for their thoughtful comments on an earlier draft of this document: Ronald Berk, John Carroll, Walter Denham, Jeremy Finn, Edward Haertel, Sylvia Johnson, Ina Mullis, Ingram Olgin, Eugene Owen, Gary Phillips, and John Tukey. However, the reviewers are not responsible for any errors that remain in the document.

APPENDIX A

PRACTICE ITEM RATING FORM

Grade Level: 8 **Judge:** ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	Rating	1st 2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
1	___	___	___	___	___	___	___	___	___	___
2	___	___	___	___	___	___	___	___	___	___
3	___	___	___	___	___	___	___	___	___	___
4	___	___	___	___	___	___	___	___	___	___
5	___	___	___	___	___	___	___	___	___	___
6	___	___	___	___	___	___	___	___	___	___
7	___	___	___	___	___	___	___	___	___	___
8	___	___	___	___	___	___	___	___	___	___
9	___	___	___	___	___	___	___	___	___	___
10	___	___	___	___	___	___	___	___	___	___
Sum =		___	___	___	___	___	___	___	___	___
$AL=\frac{Sum}{10}x100=$		___	___	___	___	___	___	___	___	___

Appendix C

Briefing Materials and Meeting Agendas

Achievement Levels Meeting

Essex Inn and Conference Center, Essex Junction, VT
August 15-17, 1990

Briefing Materials

Table of Contents

Tab	Materials
	AGENDA
A	List of Participants
B	Handbook for Judges
C	Guidelines and Nondisclosure Agreement
	ITEM SETS
D	1988 AP Calculus AB Examination 1988 AP Calculus BC Examination
E	1989 International Baccalaureate exams
F	Scholastic Aptitude Test (SAT): Math Subtests (2)
G	SAT Achievement Level I SAT Achievement Level II
H	American College Testing (ACT) Program: Math Subtest
I	1988 NAEP International math (released)
	BACKGROUND READING MATERIALS
	NCTM Standards 1990 NAEP Objectives booklet Academic Preparation for College (CEEB) Academic Preparation in Mathematics (CEEB) A Test for Our Society

Achievement Levels Meeting

Essex Inn and Conference Center, Essex Junction, VT
August 15-17, 1990

Agenda

All meetings will be held in the Governor's Mansion, located next to the Inn, but accessible inside through the lower level tunnel.

Wednesday, August 15

6:30 p.m.	Registration and Reception	Upper Level Foyer
7:30	Dinner Meeting	Upper Level
	Team Leaders' Meeting	Room M103

Thursday, August 16

8:30 a.m.	Continental Breakfast	Upper Level Foyer
9:00	Introduction to the Process of Standard Setting Review Content Descriptions of the Test and the Levels of Student Achievement	
10:15	BREAK	Upper Level Foyer
10:30	Item Security and Nondisclosure Form	
10:45	Practice Standard-Setting Exercises and Discussion	
11:45	Independent Item Ratings (First)	
12:45 p.m.	LUNCH	Lower Level Dining Room
1:30	Independent Item Ratings (First)	Upper Level
4:30	Adjourn	

Friday, August 17

8:00 a.m.	Continental Breakfast	Upper Level Foyer
8:30	Introduction to Empirical Data Independent Item Ratings (Second)	
9:45	BREAK	Upper Level Foyer
10:00	Group Discussion of Item Ratings and Preparation of Final Item Ratings	
11:30	Check-out and LUNCH	Lower Level Dining Room
12:30	Group Discussion of Item Ratings and Preparation of Final Item Ratings	
2:30	Presentation and Discussion of Achievement Levels within Each Grade Level Group	
3:00	Presentation of Grade-Level Results, Wrap-up and Future Steps	
3:30	Adjourn	

Achievement Levels Meeting

Ritz Carlton Hotel, Pentagon City
September 29 and 30, 1990

Agenda

Saturday, September 29

9:00 a.m.	Registration Continental Breakfast	2nd Floor, Foyer
10:00	Welcome Purpose of meeting	2nd Floor, Salon I
10:30	Discuss definitions	
11:00	Complete independent item ratings	
12:30 p.m.	LUNCH	Salon II
1:15	Complete independent item	Salon I
4:00	Presentation of preliminary results	

Sunday, September 30

7:30 a.m.	Continental breakfast	2nd Floor, Foyer
8:00	Discussion of grade level standards	Consulate, Gr Delegate, Gr 8 Diplomat, Gr 12
10:00	BREAK and Check-out	
10:30	Discussion of standards - total group	Salon III
11:30	LUNCH	Salon II
12:15 p.m.	Continue total group discussion	Salon II
1:30	Appropriateness ratings	
2:45	Final evaluations	

Achievement Levels Panel Meeting

The Ritz-Carlton Hotel, Alexandria, VA
November 12 and 13, 1990

Agenda

Monday, November 12

8:30 a.m.	Continental Breakfast	South Foyer
9:00	Session I Introductions Review steps in standard setting process Review all data analysis to date Review cut scores Explain process for anchoring Explain and review briefing materials	Salon II
12:00	LUNCH	
1:15 p.m.	Session II A walk-through of GRADE 4 BASIC Grade level group work	
4:30	Adjourn	

Tuesday, November 13

8:30 a.m.	Continental Breakfast	South Foyer
9:00	Session III Complete grade level group work Select items from released sets Edit and prepare grade level anchor definitions with sample items	Salon II
12:00	LUNCH	
1:15 p.m.	Session IV Across grade level sharing Discussion of consistency and coherence Final editing and preparation of anchor definitions	
3:00	Next steps in process	
3:30	Adjourn	

National Assessment Governing Board

State/Regional Site Validation Meetings
Spring, 1991

Agenda

8:00-8:30 a.m.	Registration & Continental Breakfast	
8:30-8:45 a.m.	Introductions & Welcome	NAGB Staff
8:45-9:30 a.m.	Briefing about tasks to be performed, purpose, how it will be used, etc.	Tape NAGB Staff
9:30-11:00 a.m.	Training/practice items Round 1 ratings	Tape NAGB Staff
11:00-11:15 a.m.	BREAK	
11:15-12:30 p.m.	Round 2 ratings	Tape NAGB Staff
12:30-1:30 p.m.	LUNCH	
1:30-3:00 p.m.	Within grade discussion Cross grade discussion Final ratings	Tape NAGB Staff
3:00-3:15 p.m.	BREAK	
3:15-4:15 p.m.	Probe and discussion about the original levels and definitions	Questionnaire NAGB Staff
4:15-4:45 p.m.	Evaluation and Wrap-up	NAGB Staff

Appendix D

Standard Setting Forms

Sample

Judges' Rating Form

Used

August 16-17, 1990

Essex Junction, VT

MATHEMATICS ACHIEVEMENT LEVEL SETTING FOR 1990 ASSESSMENT: AUGUST 16, 17, 1990

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
1	1									
2	2									
3	3									
4	5									
5	6									
6	7									
7	9									
8	11									
9	13									
10	14									
11	15									
12	16									
13	17									
14	18									
15	19									
16	20									
17	21									
18	22									
Sum =										

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
19	23									
20	24									
21	25									
22	26									
23	28									
24	30									
25	32									
26	33									
27	34									
28	36									
29	38									
30	39									
31	41									
32	42									
33	43									
34	44									
35	45									
36	46									
37	48									
Sum =										

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
38	49									
39	50									
40	52									
41	53									
42	54									
43	55									
44	56									
45	57									
46	58									
47	59									
48	60									
49	61									
50	62									
51	63									
52	64									
53	65									
54	66									
55	67									
56	68									
Sum =										

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
57	69									
58	70									
59	71									
60	72									
61	73									
62	74									
63	75									
64	76									
65	77									
66	78									
67	79									
68	80									
69	81									
70	82									
71	83									
72	84									
73	85									
74	86									
75	87									
Sum =										

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
76	89									
77	90									
78	91									
79	93									
80	95									
81	96									
82	97									
83	98									
84	100									
85	102									
86	103									
87	104									
88	105									
89	106									
90	107									
91	108									
92	110									
93	111									
94	112									
Sum =										

Grade Level: 4 Judge: ____________________ Booklet: __________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
95	113									
96	115									
97	117									
98	119									
99	120									
100	122									
101	124									
102	125									
103	126									
104	127									
105	129									
106	131									
107	132									
108	134									
109	136									
110	137									
111	138									
112	140									
113	142									
Sum =										

Grade Level: 4

Judge: ____________________

Booklet: __________

Item	Page	BASIC			PROFICIENT			ADVANCED		
		1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
114	143									
115	144									
116	145									
117	146									
118	147									
119	148									
120	150									
121	151									
122	152									
123	153									
124	155									
125	156									
126	157									
127	159									
128	161									
129	163									
130	165									
131	167									
132	168									
Sum =										

Grade Level: 4 Judge: ____________ Booklet: ________

		BASIC			PROFICIENT			ADVANCED		
Item	Page	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating	1st Rating	2nd Rating	Final Rating
133	170									
134	171									
135	172									
136	174									
137	176									
138	177									
139	179									
140	180									
141	181									
142	182									
143	183									
Sum =										
$AL=\frac{Sum}{143}\times100=$										

Sample

Judges' Rating Form

Used

September 29-30, 1990

Washington, DC

MATHEMATICS ACHIEVEMENT LEVEL SETTING FOR THE 1990 ASSESSMENT
(September 29, 30, 1990)

Directions

Your task in this final set of item ratings is to specify the numbers of marginally proficient, advanced, and basic students whom you would expect to answer each test item correctly at the end of the school year. Adopt the definitions of the proficient, advanced, and basic students which were originally prepared by NAGB, and which were discussed and clarified by the total group of participants in this morning's session.

In completing your item ratings, please focus attention on (1) the definitions of proficient, advanced, and basic students as you understand them, and (2) your perceptions about the difficulties of the test items when administered to marginally proficient, advanced, and basic students. You will have access to statistical data on the items. These data were obtained on a nationally representative sample of students in the spring of this year.

Note, too, we want you to provide, for each test item, your expectation of performance of marginally proficient students first, then provide your ratings of marginally advanced and marginally basic students before moving to subsequent items. Assume there are 100 marginally proficient, 100 marginally advanced, and 100 marginally basic students. Remember, the question is: How many of these students would you expect to answer each test item correctly?

Once you have completed the full rating task, please carry out the calculations needed on each page, and the final calculations on the last page.

MATHEMATICS ACHIEVEMENT LEVEL SETTING FOR THE 1990 ASSESSMENT:
(September 29, 30, 1990)

Grade Level: 8 Judge: ____________ Booklet: ________

---

*Note that your ratings should be provided in the following order: Proficient, Advanced, Basic.

Item	Page	Proficient	Advanced	Basic
1	1	____	____	____
2	2	____	____	____
3	3	____	____	____
4	5	____	____	____
5	6	____	____	____
6	7	____	____	____
7	8	____	____	____
8	9	____	____	____
9	11	____	____	____
10	13	____	____	____
11	15	____	____	____
12	16	____	____	____
13	17	____	____	____
14	18	____	____	____
15	19	____	____	____
16	20	____	____	____
17	21	____	____	____
18	22	____	____	____
19	24	____	____	____
20	25	____	____	____
21	26	____	____	____
22	27	____	____	____
23	28	____	____	____
24	29	____	____	____
25	30	____	____	____
26	31	____	____	____
27	32	____	____	____
	Sum =	____	____	____

Grade Level: 8 **Judge:** ______________ **Booklet:** __________

Item	Page	Proficient	Advanced	Basic
28	33	____	____	____
29	35	____	____	____
30	36	____	____	____
31	38	____	____	____
32	40	____	____	____
33	41	____	____	____
34	42	____	____	____
35	43	____	____	____
36	44	____	____	____
37	45	____	____	____
38	46	____	____	____
39	47	____	____	____
40	48	____	____	____
41	49	____	____	____
42	51	____	____	____
43	53	____	____	____
44	53A	____	____	____
45	54	____	____	____
46	55	____	____	____
47	56	____	____	____
48	57	____	____	____
49	58	____	____	____
50	59	____	____	____
51	60	____	____	____
52	61	____	____	____
53	62	____	____	____
54	63	____	____	____
55	64	____	____	____
56	65	____	____	____
57	66	____	____	____
58	68	____	____	____
	Sum =	____	____	____

Grade Level: 8 Judge: ____________ Booklet: __________

Item	Page	Proficient	Advanced	Basic
184	222	______	______	______
185	224	______	______	______
186	226	______	______	______
187	227	______	______	______
188	228	______	______	______
189	229	______	______	______
190	230	______	______	______
191	231	______	______	______
	Sum =	______	______	______
	Total Sum =	______	______	______
AL =	$\frac{\text{Total Sum}}{191}$ =	______	______	______

Sample

Item Appropriateness Rating Form

Used

September 29-30, 1990

Washington, DC

Item Appropriateness Ratings

The task is to indicate your perceptions about the level of appropriateness of items in the 1990 NAEP Mathematics Assessment. There are three possible ratings of item appropriateness: Low, Medium, High. To the right of each item number, circle one of these three levels to represent your views.

For each test item, the attached form contains the unique item number (be sure to match this number up to the number in the right-hand corner of the test item booklet), a brief description of what the item measures (as can be determined from the correct answer), the item difficulty, the content category and ability, and the categories of the rating scale.

Please place your name on the next page, and begin the task. At the end of the item ratings there are three open-ended questions. Your answers to these questions will be helpful to committees responsible for choosing math content for future assessments.

NAEP ACHIEVEMENT LEVEL SETTING PROJECT: GRADE 4

JUDGE: ____________________________

CODE	CONTENT DESCRIPTION	ADDITIONAL INFORMATION			ITEM APPROPRIATENESS		
N257201	90 REPRESENTS NINE TENS	0.6758	NO	CU	LOW	MEDIUM	HIGH
N257801	2753 GREATER THAN 2573, 2537, OR 2735	0.7062	NO	CU	LOW	MEDIUM	HIGH
N275401	MULT. SENTENCE FOR CIRCLES IS 5 X 3 = 15 (RATER 1)	0.7550	NO	CU	LOW	MEDIUM	HIGH
N258501	3 THIRDS ARE EQUAL TO ONE WHOLE	0.2087	NO	CU	LOW	MEDIUM	HIGH
M017701	TOTAL WEIGHT = WEIGHT OF ONE BOX X 12	0.4414	NO	CU	LOW	MEDIUM	HIGH
M018401	SIX STUDENTS SHARED EXACTLY 48 PENS	0.3767	NO	CU	LOW	MEDIUM	HIGH
M020001	642 = 4 AS FOUR TENS & 6 AS SIX HUNDREDS (RATER 1)	0.4828	NO	CU	LOW	MEDIUM	HIGH
M020101	SHADE 1/3 OF THE RECTANGLE (RATER 1)	0.1481	NO	CU	LOW	MEDIUM	HIGH
M020501	USE DOT ON NUMBER LINE TO SHOW 3/4 MARK (RATER 1)	0.2315	NO	CU	LOW	MEDIUM	HIGH
M022701	ESTIMATE: WHEN CHEN DECIDED IF HE HAD ENOUGH MONEY	0.3017	NO	CU	LOW	MEDIUM	HIGH
M022901	217 - IF 1 REPLACED BY 5: INCREASED BY 40	0.3204	NO	CU	LOW	MEDIUM	HIGH
M014701	REPLACE 5 WITH 2 TO DECREASE 5,647 TO 3,000	0.5734	NO	CU	LOW	MEDIUM	HIGH
M014801	370 IS AN EVEN NUMBER	0.3808	NO	CU	LOW	MEDIUM	HIGH
M015101	4/5 IS CLOSER THAN 2/3 TO 1	0.3296	NO	CU	LOW	MEDIUM	HIGH
N230501	459 SUBTRACTED FROM 900 IS > 400	0.4546	NO	CU	LOW	MEDIUM	HIGH
M015301	0.02 REPRESENTS THE SHADED PART OF THE FIGURE	0.5364	NO	CU	LOW	MEDIUM	HIGH

NAEP ACHIEVEMENT LEVEL SETTING PROJECT: GRADE 4

JUDGE: ______________________

CODE	CONTENT DESCRIPTION	ADDITIONAL INFORMATION			ITEM APPROPRIATENESS		
N202831	THE 3RD PICTURE SHOWS 3/4 SHADED	0.6306	NO	CU	LOW	MEDIUM	HIGH
M025831	442 > 436 IS THE TRUE STATEMENT	0.7048	NO	CU	LOW	MEDIUM	HIGH
M025931	THE CALCULATOR WITH 1376. BELONGS TO MARIA	0.5902	NO	CU	LOW	MEDIUM	HIGH
M026231	5/8 OF THE PIZZA IS STILL THERE	0.4051	NO	CU	LOW	MEDIUM	HIGH
M026331	NEED TO KNOW # PASSENGER SEATS ON PLANE	0.5489	NO	CU	LOW	MEDIUM	HIGH
M031101	93 X 76 = THE GREATEST ANSWER	0.8761	NO	CU	LOW	MEDIUM	HIGH
M031601	$3.00 IS THE TOTAL OF MONEY SHOWN	0.7217	NO	CU	LOW	MEDIUM	HIGH
M034301	TOTAL VALUE OF SYMBOLS REPRESENTED = 235 (RATER 1)	0.6357	NO	CU	LOW	MEDIUM	HIGH
M034302	USE SYMBOLS-DRAW FIG. REPRESENTING 2,041 (RATER 1)	0.6166	NO	CU	LOW	MEDIUM	HIGH
M036601	25 X 18 IS 18 MORE THAN 24 X 18	0.3200	NO	CU	LOW	MEDIUM	HIGH
N277401	JOE HAD 35 STAMPS, NOW HAS 77 AFTER BUYING 42 MORE	0.8782	NO	PK	LOW	MEDIUM	HIGH
N277601	64 - 27 = 37 (NO CALCULATOR) (RATER 1)	0.7967	NO	PK	LOW	MEDIUM	HIGH
N277602	604 - 207 = 397 (NO CALCULATOR) (RATER 1)	0.6031	NO	PK	LOW	MEDIUM	HIGH
M017401	238+462=700	0.8691	NO	PK	LOW	MEDIUM	HIGH
M018501	6 PIECES OF STRING @ 1/8" LONG = 3/4 OF A YARD	0.2385	NO	PK	LOW	MEDIUM	HIGH
N277903	SUBTRACT 65 - 7 = 58 (RATER 1)	0.7170	NO	PK	LOW	MEDIUM	HIGH

Open-Ended Comments

1. What mathematics skills and content areas do you feel were under-represented in the 1990 Assessment?

2. What mathematics skills and content areas do you feel were over-represented in the 1990 Assessment?

3. What suggestions do you have for improving the bank of items for the 1992 Assessment?

Sample

Judges' Final Rating Form

Round 3

Used

September 29-30, 1990

Washington, DC

FINAL ACHIEVEMENT LEVELS ON THE 1990 MATHEMATICS ASSESSMENT

ON THE BASIS OF (1) MY PERSONAL ITEM RATINGS; (2) DISCUSSIONS WITH MEMBERS OF MY WORK GROUP, OTHER PARTICIPANTS AT THE SAME GRADE LEVEL AS MYSELF, AND PARTICIPANTS AT THE OTHER TWO GRADE LEVELS; AND (3) THE STATISTICAL DATA I HAD AN OPPORTUNITY TO REVIEW, MY RECOMMENDED MARGINAL ACHIEVEMENT LEVELS ARE AS FOLLOWS:

________ %	________ %	________ %
BASIC	PROFICIENT	ADVANCED

HOW CONFIDENT DO YOU FEEL ABOUT THE THREE MARGINAL ACHIEVEMENT LEVELS YOU SET ABOVE?

(1 = NOT CONFIDENT 2 = SOMEWHAT CONFIDENT 3 = CONFIDENT 4 = VERY CONFIDENT)

CIRCLE ONE RATING FOR EACH ACHIEVEMENT LEVEL.

BASIC:	1	2	3	4
PROFICIENT:	1	2	3	4
ADVANCED:	1	2	3	4

HOW MIGHT THE PROCESS OF ACHIEVEMENT LEVEL SETTING HAVE BEEN REVISED TO INCREASE YOUR CONFIDENCE LEVEL? (PLEASE USE THE OPPOSITE SIDE OF THIS PAGE TO PROVIDE YOUR ANSWER.)

JUDGE: ______________________________ GRADE: __________

Sample

Judges' Rating Form

Used

Spring, 1991

Book:
Block:
Section:

MATHEMATICS ACHIEVEMENT LEVEL REPLICATION/VALIDATION PROJECT, SPRING 1991

Grade Level: 4 Judge:______________________________
Print Last Name

	BASIC		PROFICIENT		ADVANCED	
ITEM	1ST RATING	2ND RATING	1ST RATING	2ND RATING	1ST RATING	2ND RATING
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
SUM						

Book:
Block:
Section:

MATHEMATICS ACHIEVEMENT LEVEL REPLICATION/VALIDATION PROJECT, SPRING 1991

Grade Level: 4 Judge:________________________
Print Last Name

	BASIC		PROFICIENT		ADVANCED	
ITEM	1ST RATING	2ND RATING	1ST RATING	2ND RATING	1ST RATING	2ND RATING
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
SUM						

Book:
Block:
Section:

MATHEMATICS ACHIEVEMENT LEVEL REPLICATION/VALIDATION PROJECT, SPRING 1991

Grade Level: 4

Judge:____
Print Last Name

	BASIC		PROFICIENT		ADVANCED	
ITEM	1ST RATING	2ND RATING	1ST RATING	2ND RATING	1ST RATING	2ND RATING
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
SUM						

Book:
Block:
Section:

MATHEMATICS ACHIEVEMENT LEVEL REPLICATION/VALIDATION PROJECT, SPRING 1991

Grade Level: 4

Judge:______________________________
Print Last Name

	BASIC		PROFICIENT		ADVANCED	
	1ST RATING	2ND RATING	1ST RATING	2ND RATING	1ST RATING	2ND RATING
SECTION 3						
SECTION 4						
SECTION 5						
TOTAL SUM						
AL=Total Sum / Total #Items						

Sample

Judges' Final Rating Form

Round 5

Used

Spring, 1991

FINAL ACHIEVEMENT LEVELS ON THE 1990 MATHEMATICS ASSESSMENT

ON THE BASIS OF (1) MY PERSONAL ITEM RATINGS; (2) DISCUSSIONS WITH MEMBERS OF MY WORK GROUP. OTHER PARTICIPANTS AT THE SAME GRADE LEVEL AS MYSELF, AND PARTICIPANTS AT THE OTHER TWO GRADE LEVELS; AND (3) THE STATISTICAL DATA I HAD AN OPPORTUNITY TO REVIEW, MY RECOMMENDED MARGINAL ACHIEVEMENT LEVELS ARE AS FOLLOWS:

________ % BASIC

________ % PROFICIENT

________ % ADVANCED

HOW CONFIDENT DO YOU FEEL ABOUT THE THREE MARGINAL ACHIEVEMENT LEVELS YOU SET ABOVE?

(1 = NOT CONFIDENT 2 = SOMEWHAT CONFIDENT 3 = CONFIDENT 4 = VERY CONFIDENT)

CIRCLE ONE RATING FOR EACH ACHIEVEMENT LEVEL.

BASIC:	1	2	3	4
PROFICIENT:	1	2	3	4
ADVANCED:	1	2	3	4

HOW MIGHT THE PROCESS OF ACHIEVEMENT LEVEL SETTING HAVE BEEN REVISED TO INCREASE YOUR CONFIDENCE LEVEL? (PLEASE USE THE OPPOSITE SIDE OF THIS PAGE TO PROVIDE YOUR ANSWER.)

JUDGE: ______________________________ GRADE: ____________

Sample

Achievement Levels Review Form

Used

Spring, 1991

ACHIEVEMENT LEVELS REVIEW FORM
Grade 4

Name:________________________________ ID Number:____________

Directions: Please use this form to review each of the mathematics skills provided in the definition of the achievement levels for basic, proficient, and advanced students at grade four. Answer the following question for each skill listed:

> **Do you agree with the inclusion of this skill in the definition of the marginally (basic, proficient, or advanced) student?**

For each skill, **circle** the number that corresponds to your rating of the math skill for that achievement level: 1 = Yes; 2 = No; **3 = Unsure**

At the end of this form, you will be asked to list other skills that you think should have been included in the achievement levels definitions.

GRADE 4 - BASIC

> **Do you agree with the inclusion of this skill in the definition of the marginally basic student?**

RATING

Yes	No	Unsure		
1	2	3	B.1	begin to develop strategies to solve mathematical problems
1	2	3	B.2	be able to solve routine problems involving addition and subtraction, with and without the calculator
1	2	3	B.3	be able to use physical materials and pictures to help them understand and explain mathematical ideas
1	2	3	B.4	begin to develop estimative skills in measurement, numbers and computational situations
1	2	3	B.5	understand number sense and concepts related to place value
1	2	3	B.6	understand whole number operations
1	2	3	B.7	begin to develop concepts related to fractions
1	2	3	B.8	read and use simple measurement instruments
1	2	3	B.9	identify and describe simple geometric figures
1	2	3	B.10	read and use information from graphs

GRADE 4 - PROFICIENT

Do you agree with the inclusion of this skill in the definition of the marginally proficient student?

RATING

Yes	No	Unsure		
1	2	3	P.1	have an understanding of numbers and their application to life situations
1	2	3	P.2	have an understanding of measurement
1	2	3	P.3	have a knowledge of geometric figures and relationships
1	2	3	P.4	have a basic knowledge of data
1	2	3	P.5	be able to develop and apply strategies to solve a wide variety of mathematical problems
1	2	3	P.6	use patterns and relationships to analyze mathematical situations
1	2	3	P.7	relate physical materials, pictures and diagrams to mathematical ideas
1	2	3	P.8	link conceptual and procedural knowledge
1	2	3	P.9	find and use relevant information in problem solving
1	2	3	P.10	have a knowledge of numbers and concepts related to place value
1	2	3	P.11	have an understanding of whole number operations as well as a facility with whole number computation
1	2	3	P.12	be able to solve problems using a calculator
1	2	3	P.13	have the ability to use estimation skills to solve problems
1	2	3	P.14	be able to relate simple picture models to fraction symbols
1	2	3	P.15	be able to describe geometric shapes and simple attributes of these shapes
1	2	3	P.16	understand measurement concepts such as length
1	2	3	P.17	collect, interpret and display data
1	2	3	P.18	begin to develop the concept of chance
1	2	3	P.19	use simple measurement instruments

GRADE 4 - ADVANCED

Do you agree with the inclusion of this skill in the definition of the marginally advanced student?

RATING

Yes	No	Unsure		
1	2	3	A.1	be able to demonstrate flexibility in solving problems and relating knowledge to new situations
1	2	3	A.2	have an understanding of inverse relationships
1	2	3	A.3	be able to relate number concepts to more complex models and situations
1	2	3	A.4	be able to determine functional relationships from patterns
1	2	3	A.5	determine when estimation is an appropriate solution to a problem
1	2	3	A.6	read and interpret complex graphs
1	2	3	A.7	be able to use measuring instruments in non-routine ways
1	2	3	A.8	be developing concepts of decimals, symmetry, and parallelism

---

FINAL REVIEW

Look through the green 1990 NAEP Mathematics Objectives Booklet. Are there other skills on the assessment that you think should be included in the achievement levels for grade four, based on your review of items and your discussion today? Please list the numbers of the objectives (from the green booklet) under the respective achievement level.

BASIC

PROFICIENT

ADVANCED

Sample

Evaluation/Demographic Form

Used

Spring, 1991

Your ID No. ________

- National Assessment Governing Board -

Participant Survey

NAGB is interested in your views about the achievement level setting process and in the backgrounds of participants. Your answers to the questions below will be very helpful in our efforts to fully analyze the available data and to evaluate the process you went through today. Please circle the letter beside your answer to each question. Also, your ID No. above is requested to help NAGB with its analyses. Your answers will be confidential and only analyzed in conjunction with other participants who were at this meeting.

Evaluation Questions

1. What is your overall impression of the training you received today for setting achievement levels?

   a. appropriate
   b. somewhat appropriate
   c. not appropriate

2. How clear were you about NAGB's definition of the Basic student?

   a. not at all clear
   b. somewhat clear
   c. clear
   d. very clear

3. How clear were you about NAGB's definition of the Proficient student?

   a. not at all clear
   b. somewhat clear
   c. clear
   d. very clear

4. How clear were you about NAGB's definition of the Advanced student?

   a. not at all clear
   b. somewhat clear
   c. clear
   d. very clear

5. How would you judge the time allotted today to set achievement levels?

   a. not enough time
   b. too much time
   c. about the right amount of time

6. How would you judge your level of understanding of the achievement level setting process implemented today?

   a. low
   b. medium
   c. high

7. Which factors influenced the achievement levels that you set today? (Circle all choices which apply.)

   a. the definitions of basic, proficient, and advanced students
   b. the content of the items
   c. my perception of the difficulty of items
   d. actual student performance on the items
   e. persons working with the same test booklet
   f. persons working at the same grade level as myself
   g. persons working at the other grade levels
   h. other (Please specify: ______________________________)

8. What additional information and/or discussions would have been helpful to you today in setting achievement levels?

9. Do you believe that achievement levels will be useful in interpreting student performance on the 1990 NAEP Mathematics Assessment?

   a. Definitely Yes
   b. Probably Yes
   c. Unsure
   d. Probably No
   e. Definitely No

10. How successful do you believe the process was today in setting achievement levels?

   a. very successful
   b. successful
   c. somewhat successful
   d. not successful at all

11. What do you feel were the strengths of the achievement level setting process you went through today?

12. What do you feel were the weaknesses of the achievement level setting process you went through today?

Background Questions

13. Which best describes you?

    a. White
    b. Black
    c. Hispanic
    d. Asian
    e. Native American
    f. Other: ______________________

14. What is your gender?

    a. Male
    b. Female

15. Which type of organization do you represent here today?

    a. business
    b. industry
    c. school board
    d. parents
    e. educators
    f. math educators
    g. other: ______________________

16. Which best describes your current professional status?

    a. Mathematics teacher in grade 4, 8, or 12
    b. Mathematics supervisor, elementary
    c. Mathematics supervisor, secondary
    d. Mathematics supervisor, K-12
    e. School administrator
    f. Non-educator
    g. Other: ______________________

17. What type of community do you work/teach in?

    a. urban or mostly urban
    b. suburban
    c. rural or mostly rural

18. How large is the community in which you work/teach?

    a. small town
    b. large town
    c. medium city
    d. large city

* *

If you are a teacher, please answer questions 19 to 21. Others should answer question 22.

19. Approximately how many students do you teach? __________

20. What ability levels do you mostly teach?

a. average mainstream students
b. below average mainstream students
c. above average mainstream students
d. special needs students

21. How long have you been teaching?

a. 1 to 3 years
b. 4 to 10 years
c. 11 to 20 years
d. 21 years or more

* *

Only non-educators should answer question 22.

22. Which best describes the organization for whom you currently work?

a. non-profit organization
b. branch of the military
c. federal, state, local government
d. large corporation
e. small business (less than 100 employees)
f. self-employed
g. other: ______________________

* *

Thank you for taking time to complete this survey. Be sure your ID number is on page 1, and then turn in your survey.

Appendix E

Item Security Policy and Nondisclosure Form

U.S. Department of Education

Guidelines for the Release and Use of NAEP Background and Cognitive Items

The NAEP authorizing legislation, Section 406 (i) of the General Education Provisions Act (GEPA), as amended by P.L. 100-297, stipulates the following with regard to release of NAEP items:

"(4)(A) Except as provided in subparagraph (B), the public shall have access to all data, questions, and test instruments of the National Assessment.

(B)(ii) Notwithstanding any other provision of the law, the Secretary may decline to make available to the public for a period not to exceed 10 years following their initial use cognitive questions that the Secretary intends to reuse in the future."

The National Center for Education Statistics (NCES) is establishing under a delegation of authority from the Secretary of Education the following guidelines for the release and use of NAEP background and cognitive items.

1. Background items -- All NAEP background items used in collecting information on students, schools, and school staff will be available to the public.

2. Cognitive items (test items) -- All NAEP cognitive items will be divided into categories identifying their availability.

    A. NAEP cognitive items (definition) -- All cognitive items developed by NAEP become NAEP items and subject to the NAEP item release policy.

    B. General limitations on item availability

        1) Public release: Two categories of NAEP test items will be released to the public.

            a. Items more than 10 years old -- all items first used more than 10 years before the current date, and

            b. Other publicly released items -- other test items that are not intended for use in future NAEP assessments.

NCES will periodically publish NAEP test items that are available to the public.

        2) Withheld from public release: Test items withheld from public release because they are intended for use in future assessments are divided into two categories:

            a. Secured-use -- In order to provide technical assistance to the States and other users of test items, the Department is making a limited number of items withheld from the general public available under "secured use" conditions. These items will be made available only to requesters who agree to the following four conditions:

(1) They will not disclose secured-use items to anyone other than those specified on the nondisclosure agreement.

(2) They will use the same item security procedures as those used in the Trial State Assessment (or equivalent procedures acceptable to the Commissioner) in any administration of the items for assessment purposes.

(3) They will protect the rights of test takers in accordance with the professional standards in Chapter 16 of the Standards for Educational and Psychological Testing established by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (Washington, D.C., American Psychological Association, 1985).

(4) They will abide by the provision of GEPA prohibiting the use of NAEP items used in the Trial State Assessment for student, school or school district comparisons --

> "The use of National Assessment test items and test data employed in the [State] pilot program ... to rank, compare, or otherwise evaluate individual students, schools or school districts is prohibited." [Section 406 (i)(4)(C) of GEPA as amended]

b. Non-release items -- The remaining items withheld from the general public will be reserved exclusively for NAEP assessments of trends and use in future assessments.

C. Special restrictions to protect the Trial State Assessment -- All NAEP cognitive items in subject areas/grade levels covered by the Trial State Assessment will not be available to States for assessments conducted during an 18-month period before the Trial State Assessments:

1) No secured-use of NAEP 8th grade mathematics items for 18 months before the 1990 NAEP data collection, and

2) No secured-use of NAEP 4th or 8th grade mathematics items or 4th grade reading items for 18 months before the 1992 NAEP data collection.

3. "NAEP equivalent scores"

A. State assessments and other testing instruments may be linked with NAEP cognitive items to create "NAEP equivalent scores" -- scores from State and other assessment instruments which have been adjusted so that in some respects they are similar to the NAEP scale scores.

B. NAEP, NCES, and National Assessment Governing Board (NAGB) are not responsible for the degree of comparability of these "NAEP equivalent scores" and actual NAEP scores.

"NAEP equivalent scores" -- because they are not actual NAEP scores -- are not subject to the restrictive conditions imposed on NAEP cognitive items and the data generated by use of these items.

1) They are not subject to Section 406(i)(4)(C) of GEPA prohibiting student, school, and school district comparisons, and

2) They are not subject to Section 406(i)(4)(B)(i) of GEPA requiring confidentiality of individual student data.

4. The Commissioner may make exceptions to these guidelines at his discretion.

U.S. Department of Education

Item Use and Nondisclosure Agreement

I have read and understand all provisions of the U.S. Department of Education's **Guidelines for the Release and Use of NAEP Background and Cognitive Items** **(Guidelines)**.

I understand I will be working with cognitive items that are withheld from public release, and which may be used in future NAEP assessments. I agree not to disclose any such items, and further agree not to disclose the contents of any discussions conducted during these panel meetings that would reveal the specific text of these items.

____
Signature

____
Booklet Number Assigned

____
Date

[This form must be signed and submitted when receiving the item booklet.]

Appendix F

Summary of Vermont and Washington Achievement Level Setting Data

Table 1. Summary of Grade 4 Achievement Levels for the Total Item Pool

		Achievement Level					
		Basic		Proficient		Advanced	
Item Ratings	N	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1st	22	49.2	18.0	72.6	16.1	86.7	13.0
2nd	22	46.1	13.5	71.1	10.9	87.4	7.0
3rd	22	47.2	11.4	71.9	8.9	88.1	5.3
4th	11	48.5	12.6	76.0	8.6	89.2	3.7
Final	11	50.3	2.0	77.3	4.6	90.2	1.6

Table 2. Summary of Grade 8 Achievement Levels for the Total Item Pool

		Achievement Level					
		Basic		Proficient		Advanced	
Item Ratings	N	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1st	22	70.0	14.1	87.1	9.6	95.2	4.4
2nd	22	71.6	16.5	86.1	9.4	93.8	3.9
3rd	22	70.9	14.4	86.6	9.9	95.0	4.5
4th	19	68.5	12.0	84.9	8.9	93.8	4.5
Final	18	64.1	10.5	81.3	6.4	91.8	3.2

Table 3. Summary of Grade 12 Achievement Levels for Total Item Pool

		Achievement Level					
		Basic		Proficient		Advanced	
Item Ratings	N	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1st	19	53.3	16.2	81.4	8.9	95.0	3.3
2nd	19	53.4	13.6	81.7	6.9	95.0	3.2
3rd	19	54.2	11.1	82.2	5.7	95.1	3.0
4th	9	56.4	13.0	82.5	6.4	93.8	2.4
Final	9	56.4	4.7	78.0	4.0	90.8	1.4

Table 4. Summary of Grade 4 Achievement Levels for the Reduced[1] Item Pool

		Achievement Level								
		Basic			Proficient			Advanced		
Item Rating	N	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
1st	22	50.4	17.8	55.0	73.8	15.9	78.0	87.5	12.6	91.0
2nd	22	46.9	13.9	48.0	71.9	10.9	74.5	87.9	6.8	92.0
3rd	22	47.9	11.8	47.0	72.7	9.1	73.0	88.8	5.2	89.5
4th	11	49.4	12.4	48.0	76.5	9.0	79.0	89.6	4.3	89.0
Final[2]	11	50.3	2.0	50.0	77.3	4.6	75.0	90.2	1.6	90.0

[1] Excludes EST and HOTS items.

[2] Overall rating based upon the total pool of test items.

Table 5. Summary of Grade 8 Achievement Levels for the Reduced[1] Item Pool

		Achievement Level								
		Basic			Proficient			Advanced		
Item Ratings	N	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
1st	22	70.1	14.0	68.8	87.1	9.6	88.0	95.2	4.5	96.3
2nd	22	71.5	16.3	72.7	86.0	9.4	87.5	93.8	3.9	91.2
3rd	22	70.6	14.1	71.1	86.7	9.9	86.3	94.9	4.5	95.8
4th	19	68.9	13.0	69.2	85.1	9.5	86.2	93.9	5.5	94.9
Final[2]	18	64.1	10.5	60.0	81.3	6.4	80.0	91.8	3.2	92.0

[1] Excludes EST and HOTS items.

[2] Overall ratings based upon the total pool of items.

Table 6. Summary of Grade 12 Achievement Levels for the Reduced[1] Item Pool

		Achievement Level								
		Basic			Proficient			Advanced		
Item Ratings	N	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
1st	19	51.0	17.0	45.6	80.3	9.6	79.9	94.7	3.3	95.9
2nd	19	51.2	14.3	50.7	80.8	7.2	80.2	94.8	3.1	94.0
3rd	19	51.9	11.9	54.2	81.2	5.6	81.1	94.8	3.1	94.9
4th	9	54.4	13.6	51.5	81.0	7.2	82.4	93.4	2.7	92.1
Final[2]	9	56.4	4.7	55.0	78.0	4.0	80.0	90.8	1.4	90.0

[1] Excludes EST and HOTS items.

[2] Overall ratings based upon the total pool of items.

Table 7. Summary of Grade 4 Third Round Achievement Levels, Reported for Groups (N=22)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1	3rd	40.3	5.6	66.5	6.4	85.2	5.1
2	3rd	56.3	14.6	79.4	10.3	92.4	6.7
3	3rd	47.8	8.0	68.0	9.7	86.7	2.2
4	3rd	43.4	9.4	70.3	8.1	87.3	4.0
T	3rd	47.2	11.4	71.9	8.9	88.1	5.3

Table 8. Summary of Grade 8 Third Round Achievement Levels, Reported for Groups (N=22)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1	3rd	85.2	4.2	96.6	2.9	99.1	1.2
2	3rd	82.3	6.6	94.3	6.1	98.6	1.7
3	3rd	58.0	7.3	78.9	4.8	90.5	3.6
4	3rd	57.7	4.7	77.0	3.7	92.0	1.3
T	3rd	70.9	14.4	86.6	9.9	95.0	4.5

Table 9. Summary of Grade 12 Third Round Achievement Levels, Reported for Groups (N=19)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
1	3rd	66.3	8.3	84.3	2.7	95.1	0.6
2	3rd	49.7	6.2	80.6	4.3	93.7	2.1
3	3rd	45.9	12.4	77.3	5.2	93.1	3.2
4	3rd	57.2	6.6	87.2	4.9	98.5	1.9
T	3rd	54.2	11.1	82.2	5.7	95.1	3.0

Table 10. Comparison of Estimated Average Difficulties at Round 3 for Items Which Were Common to Grades 4, 8 and 12

Common Item Number	Item/Page Location Grade			Basic Grade			Proficient Grade			Advanced Grade		
	4	8	12	4	8	12	4	8	12	4	8	12
1	5-6	6-7	4-4	47	83	77	73	93	95	89	98	99
2	8-11	10-13	7-8	39	85	78	65	95	95	84	98	99
3	16-20	15-19	10-13	58	85	81	84	93	96	95	98	99
4	19-23	19-24	13-16	45	85	80	70	94	95	85	98	99
5	29-38	34-42	25-30	82	95	94	93	98	99	99	100	99
6	31-41	36-44	28-33	42	84	76	65	93	93	85	98	97
7	42-54	46-55	36-42	48	81	78	75	92	95	92	97	98
8	43-55	47-56	37-43	39	82	75	66	91	95	81	97	98
9	44-56	48-57	38-44	49	78	67	68	91	91	83	97	98
10	51-63	60-70	26-31	40	82	80	70	92	97	86	97	99
11	52-64	61-71	51-59	45	84	82	66	94	96	84	99	99
12	53-65	7-8	5-5	36	82	73	65	93	91	85	98	99
13	54-66	35-43	27-32	21	68	59	44	84	86	62	95	97
14	55-67	62-72	52-60	28	68	66	50	84	90	69	95	98
15	56-68	63-73	53-61	55	89	81	82	95	97	96	98	99
16	67-79	72-83	61-70	27	70	64	49	88	89	65	95	99
17	70-82	86-91	68-77	69	68	75	89	87	93	97	95	98
18	73-85	85-96	74-83	62	86	78	78	95	93	90	99	99
19	85-102	96-110	86-97	52	83	83	77	94	95	88	98	98
20	86-103	97-111	87-98	53	75	67	75	90	84	88	96	95
21	87-104	98-112	88-99	39	78	72	66	92	92	84	97	97
22	90-107	103-117	78-88	38	76	68	66	91	92	84	97	99
23	91-108	105-119	97-110	42	74	83	68	90	98	90	96	99
24	98-119	114-130	107-122	56	78	77	78	91	93	89	97	98
25	106-131	131-151	125-143	30	64	54	54	85	79	78	95	94
26	119-148	152-181	144-170	41	62	71	75	83	91	92	95	98
27	121-151	155-185	149-177	44	85	82	72	93	98	88	98	99

Table 10. (Continued)

Comparison of Estimated Average Difficulties at Round 3 for Items Which Were Common to Grades 4, 8 and 12

Common Item Number	Item/Page Location Grade			Basic Grade			Proficient Grade			Advanced Grade		
	4	8	12	4	8	12	4	8	12	4	8	12
28	123-153	57-66	49-55	49	85	84	74	93	97	91	98	99
29	124-155	157-186	152-181	35	73	69	64	88	91	85	96	99
30	131-167	167-200	163-196	34	65	56	59	83	86	75	94	98
31	137-176	166-199	162-195	32	66	56	60	86	83	82	955	97
32	138-177	185-224	196-242	32	69	48	60	85	82	84	96	96

Table 11. Comparison of Estimated Average Difficulties at Round 3 for Items Which Were Common to Grades 4 and 8

Common Item Number	Item/Page Location Grade		Basic Grade		Proficient Grade		Advanced Grade	
	4	8	4	8	4	8	4	8
1	6-7	8-9	55	85	82	94	95	98
2	7-9	9-11	45	88	67	95	87	99
3	9-13	11-15	42	82	71	92	88	98
4	10-14	12-16	40	80	66	92	85	98
5	24-30	28-33	18	56	39	80	62	92
6	25-32	29-35	34	72	61	90	83	97
7	38-49	38-46	52	88	76	95	96	99
8	57-69	64-74	56	88	82	95	95	99
9	58-70	65-75	38	82	69	91	85	97
10	71-83	82-93	38	74	65	88	87	96
11	78-91	89-100	48	85	73	94	89	98
12	79-93	90-102	45	83	75	94	90	98
13	80-95	91-104	41	85	66	94	85	98
14	96-115	111-126	23	61	48	81	69	92
15	99-120	116-132	20	80	43	92	67	98
16	105-129	128-147	35	71	63	87	82	96
17	107-132	132-152	34	73	55	86	73	95
18	108-134	135-158	32	70	55	86	76	91
19	109-136	136-160	21	65	47	86	74	95
20	112-140	140-164	16	62	44	88	69	95
21	114-143	145-172	27	72	56	89	80	97
22	128-161	163-194	36	78	71	90	91	96
23	132-168	170-203	50	83	76	93	93	98
24	133-170	172-207	43	78	71	91	89	97
25	139-179	186-226	58	85	82	95	96	98
26	140-180	187-227	31	66	63	86	83	95
27	143-183	191-231	27	69	56	87	81	96

Table 12. Comparison of Estimated Average Difficulties at Round 3 for Randomly Selected (50%) Common Items to Grades 8 and 12

Common Item Number	Item/Page Location Grade		Basic Grade		Proficient Grade		Advanced Grade	
	8	12	8	12	8	12	8	12
1	14-18	9-12	59	46	79	78	92	94
2	21-26	15-18	79	68	92	90	97	98
3	23-28	17-20	76	65	91	88	97	97
4	25-30	19-22	84	78	94	96	98	99
5	27-32	21-24	62	50	84	81	93	96
6	49-58	39-45	76	76	90	93	96	98
7	51-60	41-47	60	56	82	83	91	95
8	53-62	43-49	85	86	93	96	98	99
9	55-64	45-51	82	78	93	94	97	98
10	58-68	47-53	78	80	88	95	96	98
11	66-76	56-65	53	29	75	63	90	84
12	68-78	31-37	70	62	87	87	95	97
13	73-84	62-71	62	55	81	80	93	95
14	75-86	64-73	67	73	86	90	94	99
15	77-88	66-75	63	58	81	85	92	96
16	86-97	75-84	68	58	87	89	95	98
17	94-108	84-95	88	85	96	98	99	99
18	99-113	89-100	69	74	87	91	96	98
19	101-115	91-102	74	78	89	92	96	98
20	104-118	79-89	48	38	73	74	90	96
21	109-124	103-118	49	45	74	73	89	94
22	115-131	108-123	56	52	80	79	91	96
23	118-136	111-127	69	66	88	87	96	98
24	120-138	113-129	73	60	89	90	96	99
25	125-143	116-132	72	64	85	90	95	99
26	129-149	121-139	64	55	85	89	96	99
27	138-162	120-138	52	39	78	78	90	95

Table 12. (Continued)

Comparison of Estimated Average Difficulties at Round 3 for Randomly Selected (50%) Common Items to Grades 8 and 12

Common Item Number	Item/Page Location Grade		Basic Grade		Proficient Grade		Advanced Grade	
	8	12	8	12	8	12	8	12
28	143-168	136-160	54	34	76	68	89	91
29	146-173	139-165	72	66	89	89	96	97
30	152-180	143-169	62	46	83	80	95	93
31	156-185	150-178	74	67	87	90	95	97
32	159-188	145-172	68	58	87	84	96	97
33	161-191	154-183	58	40	79	70	90	92
34	164-196	159-191	56	51	77	79	92	94
35	169-202	177-213	46	24	69	74	86	95
36	173-208	168-202	60	39	79	78	90	97
37	177-213	174-209	50	37	74	68	89	92
38	181-218	183-220	81	79	94	94	98	99
39	189-229	202-250	38	22	62	63	76	84

Table 13. Performance of the Average Student in the 1990 National Sample on Common Math Items

Grade	Number of Items	Average Item Performance[1] Grade 4	8	12
4, 8, 12	32	.42	.62	.76
4, 8	27	.31	.61	--
8, 12[2]	39	--	.47	.61

[1] Average Item Performance (Complete Pool of Items)

$\overline{x}_4 = .48$, $\overline{x}_8 = .53$, $\overline{x}_{12} = .55$

[2] A 50% random sample of items was selected.

Table 14. Summary of Average Item Performance and Achievement Levels on the Common Items After the Third Set of Ratings

	Empirical Data			Judgmental Data								
	Average Item p-value			Basic			Proficient			Advanced		
	Grade			Grade			Grade			Grade		
Number of Items	4	8	12	4	8	12	4	8	12	4	8	12
32	.42	.62	.76	.44	.78	.73	.69	.91	.92	.85	.97	.98
27	.31	.61	--	.37	.76	--	.64	.90	--	.83	.97	--
39	--	.47	.61	--	.66	.57	--	.84	.84	--	.93	.96

Table 15. Summary of Judges' Five Sets of Achievement Levels (Grade 4, 22 Judges)

	Basic					Proficient					Advanced				
ID	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0405	68	50	47	--	--	89	77	75	--	--	97	91	89	--	--
0401	54	45	45	48	48	79	73	73	88	76	93	90	90	96	90
0403	59	19	29	29	50	88	54	59	68	75	97	78	82	89	90
0407	18	34	35	--	--	37	58	58	--	--	53	75	77	--	--
0422	73	52	47	55	55	87	75	73	84	85	92	89	89	94	90
0415	28	39	40	37	50	49	67	67	64	75	71	87	87	82	90
0419	49	49	45	--	--	77	76	72	--	--	90	91	89	--	--
0412	53	54	55	53	53	76	77	77	88	88	90	91	91	94	95
0404	64	65	67	--	--	87	87	87	--	--	97	97	97	--	--
0416	59	54	53	--	--	83	79	78	--	--	97	94	93	--	--
0414	69	81	79	--	--	91	97	95	--	--	99	100	100	--	--
0424	34	38	45	--	--	61	67	71	--	--	78	84	87	--	--
0408	15	39	45	67	50	34	60	66	80	76	53	80	83	89	89
0423	41	52	52	--	--	72	77	77	--	--	85	89	89	--	--
0410	27	30	38	32	50	66	60	64	65	75	89	85	86	88	90
0409	65	62	59	60	50	78	76	75	79	75	89	88	88	91	90
0411	60	52	49	--	--	83	79	76	--	--	96	94	91	--	--
0417	33	37	38	58	50	69	71	72	79	75	93	93	93	86	90
0425	47	41	46	46	47	65	58	67	70	72	77	71	80	87	88
0402	76	55	54	--	--	85	78	78	--	--	93	89	89	--	--
0421	57	36	37	--	--	77	60	61	--	--	91	83	83	--	--
0413	32	32	35	48	50	61	58	62	68	76	84	83	85	85	90
Mean	49	46	47	49	50	72	71	72	76	77	87	87	88	89	90
SD	18	14	11	12	2	16	11	9	9	5	13	7	5	4	2

Table 16. Summary of Judges' Five Sets of Achievement Levels (Grade 8, 22 Judges)

	Basic					Proficient					Advanced				
ID	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0808	87	89	86	--	--	92	90	92	--	--	97	91	97	--	--
0802	60	61	61	53	60	77	78	78	75	76	91	92	92	88	90
0811	74	84	84	80	60	99	100	100	96	75	99	100	100	98	90
0815	90	87	86	75	66	100	89	98	94	83	100	90	100	99	93
0827	72	80	78	--	--	97	97	97	--	--	100	100	100	--	--
0821	66	73	82	85	85	89	97	98	97	92	99	100	100	100	97
0806	81	81	84	67	60	98	98	98	91	80	99	99	99	97	90
0820	77	90	77	61	60	89	90	89	83	80	96	90	97	96	93
0812	91	93	93	92	88	100	100	100	98	95	100	100	100	100	99
0816	72	74	76	61	60	82	84	86	72	79	93	95	97	86	91
0825	75	88	85	78	65	84	94	94	87	83	92	96	98	94	90
0803	53	53	53	73	60	70	73	73	83	78	87	91	90	95	92
0828	40	44	44	47	50	87	86	86	80	80	95	95	95	93	90
0810	63	88	59	45	50	76	89	72	62	70	88	90	84	78	85
0822	93	90	91	85	80	99	91	99	94	92	100	91	100	97	94
0823	65	66	66	73	64	81	82	82	85	84	91	91	91	94	94
0807	59	59	58	57	55	79	79	77	74	75	91	91	90	88	88
0801	85	70	65	60	65	92	82	78	63	82	100	97	92	77	92
0826	66	42	58	64	60	92	71	82	87	80	98	91	94	98	92
0824	53	55	58	64	--	71	72	75	83	--	90	91	92	97	--
0805	63	52	54	--	--	86	76	76	--	--	96	92	92	--	--
0809	55	56	60	69	65	76	76	77	86	80	90	90	91	95	92
Mean	70	72	71	69	64	87	86	87	85	81	95	94	95	94	92
SD	14	17	14	12	11	10	9	10	9	6	4	4	5	7	3

Table 17. Summary of Judges' Five Sets of Achievement Levels (Grade 12, 19 Judges)

	Basic					Proficient					Advanced				
ID	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1213	37	38	47	--	--	86	91	92	--	--	99	100	100	--	--
1215	44	54	58	--	--	74	84	88	--	--	90	97	99	--	--
1221	46	45	45	53	60	81	82	83	75	75	97	97	97	94	90
1208	26	26	30	--	--	62	67	72	--	--	97	98	96	--	--
1212	50	50	50	55	53	75	75	75	84	80	91	91	91	93	91
1202	36	40	42	--	--	80	79	80	--	--	94	94	95	--	--
1223	59	56	53	45	58	96	94	87	84	80	100	99	96	92	90
1203	76	73	64	--	--	86	83	81	--	--	93	91	91	--	--
1210	45	45	46	55	55	78	78	78	80	75	92	92	92	93	90
1219	76	74	72	--	--	90	87	86	--	--	97	96	95	--	--
1205	59	56	58	43	50	80	79	82	75	75	94	93	95	92	90
1222	39	44	44	46	52	67	70	71	72	72	88	89	89	90	92
1204	71	68	65	--	--	92	91	91	--	--	98	99	100	--	--
1217	53	58	57	58	60	80	82	82	86	80	98	98	98	96	90
1220	71	57	65	--	--	89	77	84	--	--	96	92	95	--	--
1209	39	39	46	--	--	79	80	81	--	--	94	94	94	--	--
1201	80	77	74	72	65	90	88	87	86	85	98	97	95	94	94
1207	41	50	55	81	55	71	79	81	94	80	91	94	94	99	90
1206	65	63	58	--	--	89	84	80	--	--	97	93	92	--	--
Mean	53	53	54	56	56	81	82	82	83	78	95	95	95	94	91
SD	16	14	11	13	5	9	7	6	6	4	3	3	3	2	1

Table 18. Final 1990 NAEP Total Item Pool Mathematics Assessment Achievement Levels[1, 2]

	Achievement Level		
Grade	Basic	Proficient	Advanced
4	50	77	90
8	64	81	92
12	56	78	91

[1] Achievement levels across grade levels are not easily compared because the content specifications for items at the three grade levels are different.

[2] Based on the final set of ratings (38 judges).

Table 19. Descriptive Statistics on the Final Total Item Pool Mathematics Achievement Levels

		Achievement Level					
		Basic		Proficient		Advanced	
Grade	Judges	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
4	11	50.3	2.0	77.3	4.6	90.2	1.6
8	18	64.1	10.5	81.3	6.4	91.8	3.2
12	9	56.4	4.7	78.0	4.0	90.8	1.4

Table 20. Summary of Confidence Levels of Judges in Setting Final Achievement Levels

			Confidence Level			
Grade	Judges	Level	Not Confident	Somewhat Confident	Confident	Very Confident
4	11	Basic	0	1	4	6
		Proficient	0	0	6	5
		Advanced	0	0	4	7
8	18	Basic	0	1	10	7
		Proficient	0	0	6	12
		Advanced	0	0	6	12
12	9	Basic	0	2	5	2
		Proficient	0	0	3	6
		Advanced	0	0	3	6

Table 21. Return Rates of Judges to the Washington Meeting

	Number of Judges		
Grade	Vermont	Washington	% Return
4	22	11	50
8	22	19	86
12	19	9	47
Total	63	39	62

Table 22. Comparison of the Demographic Composition of Judges at the Vermont and Washington Meetings

	Vermont		Washington	
	N	%	N	%
Educator	45	71	32	84
Non-Educator	18	29	6	16
Ethnicity				
White	52	83	29	76
Black	8	13	6	16
Hispanic	1	1	1	2
Asian	1	1	1	2
Native American	1	1	1	2
Gender				
Male	30	48	16	42
Female	33	52	22	58

Table 23. Comparison of 3rd Set of (Vermont) Ratings for Judges "Not Present" and "Present" at the Washington Meeting

Grade	Achievement Level	Not Present in Washington	Present in Washington
		(N=11)	(N=11)
4	Basic	51.2	43.4
	Proficient	75.3	68.6
	Advanced	89.5	86.7
		(N= 3)	(N=19)
8	Basic	72.6	70.6
	Proficient	88.3	86.5
	Advanced	96.3	94.9
		(N=11)	(N= 8)
12	Basic	55.0	53.0
	Proficient	83.4	80.7
	Advanced	95.7	94.2

Explanation of the Adjustments in Tables 24, 25, 26

Tables 24 to 26 were used by 12 judges in preparing skill descriptions of the marginally basic, proficient, and advanced students. The numbers in tables 24 to 26 are the (adjusted) averages of the total group of judges' achievement levels at the item level from round four. Of course, these 12 judges should have used the item statistics based on the final (fifth) round of ratings, but these ratings were not provided at the item level. Therefore, the item ratings at the fourth round were used to reflect the final item ratings, but they were adjusted to highlight changes in the overall achievement levels between the fourth and final ratings. The adjustments based upon (mean) achievement levels in Tables 1, 3, and 5 are shown below:

Level	4th Round	Final Round	Adjustment
Grade 4			
Basic	49.4%	50.5%	+1%
Proficient	76.5%	77.3%	+1%
Advanced	89.6%	90.2%	+1%
Grade 8			
Basic	68.9%	64.1%	-5%
Proficient	85.1%	81.3%	-4%
Advanced	93.9%	91.8%	-3%
Grade 12			
Basic	54.4%	56.4%	+2%
Proficient	81.0%	78.0%	-3%
Advanced	93.4%	90.8%	-3%

Table 24. Average (Adjusted)[1] Grade 4 Item Achievement Levels

Content Category	Item	Page	Achievement Level		
			Basic	Proficient	Advanced
Numbers and Operations	1	1	68	91	97
	2	2	76	93	98
	3	3	69	90	94
	4	5	59	82	94
	5	6	51	77	93
	6	7	60	82	95
	7	9	55	81	93
	8	11	45	72	90
	9	13	51	76	92
	10	14	49	75	91
	11	15	50	74	92
	12	16	68	89	97
	13	17	39	71	86
	14	18	55	80	94
	15	19	47	77	90
	16	20	64	87	96
	17	21	69	87	95
	18	22	58	81	94
	19	23	--	--	--
	20	24	--	--	--
	21	25	--	--	--
	22	26	--	--	--
	23	28	--	--	--
	24	30	--	--	--
	25	32	--	--	--
	26	33	79	95	98
	27	34	77	96	99
	28	36	63	89	98
	29	38	76	94	99
	30	39	73	94	98
	31	41	51	80	91
	32	42	54	84	95
	33	43	61	87	98
	34	44	66	90	98
	35	45	59	84	95
	36	46	60	82	94
	37	48	59	82	94
	38	49	53	80	94
	39	50	70	91	97
	40	52	57	84	94
	41	53	55	78	90
	42	54	--	--	--
	43	55	--	--	--
	44	56	--	--	--
	45	57	--	--	--
	46	58	--	--	--
	47	59	--	--	--
	48	60	--	--	--
	49	61	50	82	93
	50	62	52	79	91
	51	63	48	78	95
	52	64	48	82	94
	53	65	43	75	89
	54	66	30	57	75
	55	67	36	66	83
	56	68	55	84	94
	57	69	58	86	96
	58	70	49	80	91

Table 24. (Continued)

Average (Adjusted) Grade 4 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	 Proficient	 Advanced
	59	71	50	80	93
	60	72	41	76	89
	61	73	44	78	92
	62	74	51	81	93
	63	75	55	83	95
	64	76	47	79	94
	65	77	47	78	93
	66	78	27	59	74
	67	79	--	--	--
Measurement	68	80	72	92	99
	69	81	54	84	95
	70	82	70	90	98
	71	83	44	78	90
	72	84	63	89	97
	73	85	--[2]	--	--
	74	86	--	--	--
	75	87	--	--	--
	76	89	42	77	90
	77	90	42	74	88
	78	91	50	84	93
	79	93	40	77	92
	80	95	39	71	86
	81	96	62	88	98
	82	97	47	79	92
	83	98	42	72	89
	84	100	41	69	85
	85	102	--	--	--
	86	103	--	--	--
	87	104	--	--	--
	88	105	--	--	--
	89	106	38	72	88
	90	107	35	71	87
	91	108	40	74	91
	92	110	45	77	92
	93	111	62	87	97
	94	112	41	76	91
	95	113	--	--	--
	96	115	--	--	--
Geometry	97	117	56	85	95
	98	119	55	83	93
	99	120	32	59	70
	100	122	28	57	74
	101	124	45	73	87
	102	125	42	77	91
	103	126	73	92	98
	104	127	40	66	81
	105	129	41	67	86
	106	131	29	60	77
	107	132	36	66	79
	108	134	37	57	72
	109	136	26	53	75
	110	137	31	58	76
	111	138	--	--	--
	112	140	--	--	--

Table 24. (Continued)
Average (Adjusted) Grade 4 Item Achievement Levels

Content Category	Item	Page	Achievement Level		
			Basic	Proficient	Advanced
Data Analysis, Statistics, and Probability	113	142	34	67	80
	114	143	34	66	83
	115	144	53	82	93
	116	145	51	77	93
	117	146	50	79	89
	118	147	47	78	89
	119	148	50	79	90
	120	150	57	86	96
	121	151	--	--	--
	122	152	--	--	--
	123	153	--	--	--
	124	155	33	70	87
	125	156	--	--	--
	126	157	--	--	--
	127	159	--	--	--
	128	161	--	--	--
Algebra and Functions	129	163	79	97	99
	130	165	60	90	97
	131	167	34	63	77
	132	168	48	82	93
	133	170	49	80	91
	134	171	61	87	96
	135	172	42	76	90
	136	174	26	59	73
	137	176	28	67	84
	138	177	33	63	78
	139	179	62	88	95
	140	180	30	64	79
	141	181	52	80	93
	142	182	61	86	95
	143	183	--	--	--

---

[1] Added 1% to Basic, Proficient, and Advanced.

[2] Data on HOTS and EST items which were deleted are not included.

Table 25. Average (Adjusted)[1] Grade 8 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	Proficient	Advanced
Numbers and Operations	1	1	76	88	95
	2	2	72	85	94
	3	3	73	88	95
	4	5	74	89	94
	5	6	67	84	94
	6	7	78	88	95
	7	8	76	87	94
	8	9	77	89	95
	9	1	82	91	96
	10	13	76	88	94
	11	15	74	89	94
	12	16	76	88	94
	13	17	66	83	92
	14	18	53	78	89
	15	19	80	90	95
	16	20	47	71	85
	17	21	59	80	91
	18	22	49	74	87
	19	24	--[2]	--	--
	20	25	--	--	--
	21	26	--	--	--
	22	27	--	--	--
	23	28	--	--	--
	24	29	--	--	--
	25	30	--	--	--
	26	31	--	--	--
	27	32	--	--	--
	28	33	--	--	--
	29	35	--	--	--
	30	36	84	91	95
	31	38	84	91	95
	32	40	78	88	95
	33	41	70	86	93
	34	42	90	93	96
	35	43	64	80	91
	36	44	74	85	94
	37	45	60	79	90
	38	46	76	89	94
	39	47	71	85	92
	40	48	80	89	94
	41	49	52	75	89
	42	51	61	80	91
	43	53	64	81	91
	44	53A	48	69	86
	45	54	49	72	87
	46	55	--	--	--
	47	56	--	--	--
	48	57	--	--	--
	49	58	--	--	--
	50	59	--	--	--
	51	60	--	--	--
	52	61	--	--	--
	53	62	--	--	--
	54	63	--	--	--
	55	64	--	--	--
	56	65	--	--	--
	57	66	--	--	--
	58	68	--	--	--

Table 25. (Continued)

Average (Adjusted) Grade 8 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	Achievement Level Proficient	Achievement Level Advanced
	59	69	--	--	--
	60	70	75	87	93
	61	71	73	85	94
	62	72	54	74	88
	63	73	80	89	94
	64	74	79	90	95
	65	75	73	86	94
	66	76	45	67	82
	67	77	68	83	92
	68	78	59	79	90
	69	79	45	71	87
	70	80	60	80	90
	71	81	44	67	84
	72	83	--	--	--
	73	84	--	--	--
	74	85	--	--	--
	75	86	--	--	--
	76	87	--	--	--
	77	88	--	--	--
Measurement	78	89	83	91	95
	79	90	68	84	92
	80	91	86	92	95
	81	92	48	70	86
	82	93	64	79	91
	83	94	66	82	92
	84	95	79	90	95
	85	96	--	--	--
	86	97	--	--	--
	87	98	82	90	96
	88	99	73	86	94
	89	100	76	87	94
	90	102	75	87	93
	91	104	72	87	95
	92	105	60	78	89
	93	107	71	86	94
	94	108	85	92	96
	95	109	81	91	96
	96	110	--	--	--
	97	111	--	--	--
	98	112	--	--	--
	99	113	--	--	--
	100	114	--	--	--
	101	115	--	--	--
	102	116	--	--	--
	103	117	70	84	94
	104	118	43	65	82
	105	119	72	85	93
	106	121	59	78	89
	107	122	37	64	82
	108	123	--	--	--
	109	124	--	--	--
	110	125	--	--	--
	111	126	--	--	--
Geometry	112	128	69	84	93
	113	129	69	83	92
	114	130	74	86	93
	115	131	51	75	8

Table 25. (Continued)

Average (Adjusted) Grade 8 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	 Proficient	 Advanced
	116	132	72	87	94
	117	134	55	76	88
	118	136	63	81	91
	119	137	68	83	93
	120	138	60	81	91
	121	139	51	78	89
	122	140	73	88	94
	123	141	71	84	93
	124	142	56	76	86
	125	143	--	--	--
	126	144	45	71	86
	127	145	50	73	89
	128	147	69	83	92
	129	149	62	82	91
	130	150	65	83	92
	131	151	62	79	90
	132	152	65	80	91
	133	154	55	75	88
	134	156	56	76	87
	135	158	56	75	85
	136	160	58	77	87
	137	161	53	75	88
	138	162	43	70	86
	139	163	--	--	--
	140	164	--	--	--
Data Analysis, Statistics, and Probability	141	166	72	85	92
	142	167	77	89	94
	143	168	44	69	81
	144	170	60	79	90
	145	172	72	84	92
	146	173	68	83	91
	147	174	61	79	90
	148	175	57	77	89
	149	176	62	80	92
	150	178	80	90	95
	151	179	70	87	94
	152	180	49	71	84
	153	181	78	89	95
	154	183	60	80	91
	155	184	--	--	--
	156	185	--	--	--
	157	186	66	85	94
	158	187	48	72	85
	159	188	60	79	91
	160	189	42	67	81
	161	191	43	68	80
	162	193	--	--	--
	163	194	--	--	--
	164	196	--	--	--
Algebra and Functions	165	198	52	76	88
	166	199	51	80	92
	167	200	60	81	92
	168	201	52	75	88
	169	202	42	69	84
	170	203	75	86	94
	171	205	49	77	90
	172	207	72	85	93

Table 25. (Continued)

Average (Adjusted) Grade 8 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	Proficient	Advanced
	173	208	44	67	83
	174	209	42	68	83
	175	211	79	90	95
	176	212	73	87	95
	177	213	--	--	--
	178	214	78	89	96
	179	216	46	74	87
	180	217	52	79	91
	181	218	80	90	95
	182	219	41	69	86
	183	220	57	78	90
	184	222	60	79	90
	185	224	62	79	91
	186	226	84	91	96
	187	227	63	82	93
	188	228	41	71	86
	189	229	34	65	81
	190	230	58	80	91
	191	231	--	--	--

---

[1] Dropped 5% from Basic, dropped 4% from Proficient, and dropped 3% from Advanced.

[2] Data on HOTS and EST items which were deleted are not included.

Table 26. Average (Adjusted)[1] Grade 12 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	 Proficient	 Advanced
Numbers and Operations	1	1	78	86	95
	2	2	75	88	96
	3	3	60	82	93
	4	4	80	90	96
	5	5	81	90	96
	6	6	70	87	96
	7	8	83	92	97
	8	10	49	77	90
	9	12	50	75	87
	10	13	83	94	96
	11	14	41	74	89
	12	15	80	92	96
	13	16	--[2]	--	--
	14	17	--	--	--
	15	18	--	--	--
	16	19	--	--	--
	17	20	--	--	--
	18	21	--	--	--
	19	22	--	--	--
	20	23	--	--	--
	21	24	--	--	--
	22	25	89	94	96
	23	26	90	95	96
	24	28	84	92	96
	25	30	93	96	96
	26	31	85	93	96
	27	32	67	86	93
	28	33	88	95	96
	29	34	77	92	95
	30	36	65	82	93
	31	37	63	81	92
	32	38	80	92	96
	33	39	50	79	90
	34	40	62	80	91
	35	41	33	64	77
	36	42	--	--	--
	37	43	--	--	--
	38	44	--	--	--
	39	45	--	--	--
	40	46	--	--	--
	41	47	--	--	--
	42	48	--	--	--
	43	49	--	--	--
	44	50	--	--	--
	45	51	--	--	--
	46	52	--	--	--
	47	53	--	--	--
	48	54	--	--	--
	49	55	--	--	--
	50	57	77	88	95
	51	59	85	93	97
	52	60	61	83	93
	53	61	89	96	97
	54	62	40	65	83
	55	63	46	71	88
	56	65	34	59	83
	57	66	84	92	96
	58	67	52	80	92

Table 26. (Continued)

Average (Adjusted) Grade 12 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	 Proficient	 Advanced
	59	68	75	88	94
	60	69	61	82	92
	61	70	--	--	--
	62	71	--	--	--
	63	72	--	--	--
	64	73	--	--	--
	65	74	--	--	--
	66	75	--	--	--
Measurement	67	76	69	87	95
	68	77	93	96	97
	69	78	54	78	90
	70	79	63	82	94
	71	80	61	81	94
	72	81	66	81	93
	73	82	58	82	95
	74	83	--	--	--
	75	84	--	--	--
	76	85	--	--	--
	77	87	47	75	90
	78	88	75	87	96
	79	89	41	71	86
	80	90	88	94	96
	81	91	37	73	89
	82	93	61	80	93
	83	94	81	92	97
	84	95	90	95	97
	85	96	88	94	96
	86	97	--	--	--
	87	98	--	--	--
	88	99	--	--	--
	89	100	--	--	--
	90	101	--	--	--
	91	102	--	--	--
	92	103	--	--	--
	93	104	--	--	--
	94	106	--	--	--
	95	108	85	91	96
	96	109	56	81	93
	97	110	87	94	96
	98	112	31	64	84
	99	113	16	49	74
	100	115	46	72	88
	101	116	40	65	81
	102	117	--	--	--
	103	118	--	--	--
	104	119	--	--	--
Geometry	105	120	80	92	96
	106	121	84	93	96
	107	122	83	89	94
	108	123	61	83	93
	109	124	45	75	90
	110	126	65	86	94
	111	127	75	87	96
	112	128	76	88	94
	113	129	68	88	95
	114	130	61	83	94
	115	131	84	92	95

Table 26. (Continued)

Average (Adjusted) Grade 12 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	Achievement Level Proficient	Achievement Level Advanced
	116	132	--	--	--
	117	133	--	--	--
	118	135	71	88	95
	119	137	44	74	89
	120	138	47	78	92
	121	139	75	90	96
	122	140	50	80	93
	123	141	63	84	94
	124	142	58	80	93
	125	143	63	83	93
	126	144	21	60	81
	127	146	51	77	92
	128	147	43	73	88
	129	149	21	63	84
	130	151	19	52	75
	131	152	--	--	--
	132	153	--	--	--
Data Analysis, Statistics, and Probability	133	155	89	92	96
	134	156	74	87	95
	135	158	83	93	96
	136	160	33	63	85
	137	162	70	82	94
	138	164	59	79	91
	139	165	63	81	93
	140	166	72	87	95
	141	167	59	76	90
	142	168	82	89	96
	143	169	46	75	85
	144	170	90	93	96
	145	172	69	83	93
	146	173	79	91	95
	147	174	56	79	92
	148	176	79	90	96
	149	177	--	--	--
	150	178	--	--	--
	151	179	70	84	93
	152	181	83	93	97
	153	182	53	77	92
	154	183	39	65	85
	155	185	20	49	73
	156	186	19	44	62
	157	188	--	--	--
	158	189	--	--	--
	159	191	--	--	--
Algebra and Functions	160	193	31	66	84
	161	194	27	61	83
	162	195	74	85	95
	163	196	68	84	94
	164	197	58	83	93
	165	198	15	49	76
	166	199	9	40	76
	167	200	65	86	95
	168	202	40	73	90
	169	203	25	58	83
	170	205	43	73	91
	171	206	12	44	79
	172	207	44	71	88

Table 26. (Continued)

Average (Adjusted) Grade 12 Item Achievement Levels

Content Category	Item	Page	Achievement Level Basic	 Proficient	 Advanced
	173	208	10	48	75
	174	209	--	--	--
	175	210	69	91	96
	176	211	50	80	87
	177	213	38	76	90
	178	214	41	75	91
	179	215	71	90	96
	180	216	19	55	84
	181	217	63	85	94
	182	218	9	50	75
	183	220	93	95	96
	184	221	14	51	77
	185	222	14	63	86
	186	224	30	74	89
	187	225	5	48	79
	188	227	--	--	--
	189	229	--	--	--
	190	231	--	--	--
	191	233	--	--	--
	192	235	--	--	--
	193	237	47	75	90
	194	238	22	58	81
	195	240	21	57	85
	196	242	55	79	91
	197	244	13	55	80
	198	245	29	67	86
	199	247	34	69	86
	200	248	32	71	88
	201	249	10	51	80
	202	250	34	63	82
	203	251	5	41	68

---

[1] Added 2% to Basic; dropped 3% from Proficient, and 3% from Advanced.

[2] Data for HOTS and EST items which were deleted are not included.

Table 27. Summary of Judges' Five Sets of Achievement Levels for the Reduced[1] Item Pool (Grade 4, 22 Judges)

		Basic					Proficient					Advanced				
ID	ED[2]	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0405	1	68	50	47	--	--	90	77	75	--	--	98	91	89	--	--
0401	2	54	46	46	48	48	80	74	74	89	76	94	91	91	96	90
0403	1	60	19	28	29	50	89	54	59	69	75	98	78	82	89	90
0407	1	19	34	35	--	--	38	60	59	--	--	56	77	78	--	--
0422	1	75	53	48	57	55	88	75	73	86	85	93	89	90	95	90
0415	1	28	39	40	38	50	50	68	68	65	75	71	87	88	83	90
0419	2	50	51	45	--	--	78	78	73	--	--	92	92	90	--	--
0412	1	56	56	57	54	53	78	78	79	88	88	90	91	92	95	95
0404	2	65	66	69	--	--	88	88	89	--	--	98	98	98	--	--
0416	1	61	56	55	--	--	84	80	79	--	--	97	94	93	--	--
0414	2	69	82	80	--	--	91	98	96	--	--	99	100	100	--	--
0424	1	36	40	47	--	--	63	68	72	--	--	79	84	88	--	--
0408	1	17	39	45	68	50	36	60	66	80	76	55	81	84	89	89
0423	1	42	53	52	--	--	74	78	78	--	--	86	90	90	--	--
0410	2	29	30	38	32	50	69	61	66	66	75	90	86	87	89	90
0409	1	65	62	59	63	50	78	76	75	80	75	89	89	88	91	90
0411	2	63	53	49	--	--	85	80	77	--	--	97	94	92	--	--
0417	1	34	37	39	58	50	71	72	73	79	75	94	93	93	86	90
0425	1	49	43	48	48	47	66	60	68	71	72	78	72	81	88	88
0402	1	76	55	54	--	--	86	78	78	--	--	93	90	90	--	--
0421	2	58	36	37	--	--	78	61	61	--	--	92	84	84	--	--
0413	1	34	31	35	48	50	63	58	62	68	76	85	83	85	85	90
Mean		50.4	46.9	47.9	49.4	50.3	73.8	71.9	72.7	76.5	77.3	87.5	87.9	88.8	89.6	90.2
SD		17.8	13.9	11.8	12.4	2.0	15.9	10.9	9.1	9.0	4.6	12.6	6.8	5.2	4.3	1.6
Median		55.0	48.0	47.0	48.0	50.0	78.0	74.5	73.0	79.0	75.0	91.0	92.0	89.5	89.0	90.0

[2] Excludes EST and HOTS Items.

[1] Educator: 1=Yes; 2=No

Table 28. Summary of Judges' Five Sets of Achievement Levels for the Reduced[1] Item Pool (Grade 8, 22 Judges)

		Basic					Proficient					Advanced				
ID	ED[2]	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
0808	1	86	89	85	--	--	91	91	92	--	--	96	91	96	--	--
0802	1	58	60	60	53	60	75	77	77	75	76	90	92	92	87	90
0811	1	76	84	84	81	60	99	99	99	96	75	99	100	100	98	90
0815	1	91	87	86	75	66	99	89	97	94	83	100	90	99	99	93
0827	2	69	79	77	--	--	97	97	97	--	--	100	100	100	--	--
0821	1	64	72	81	84	85	88	97	99	97	92	99	100	100	100	97
0806	1	81	79	83	66	60	98	98	98	91	80	100	99	99	97	90
0820	1	76	90	76	61	60	89	90	89	82	80	96	90	97	95	93
0812	1	91	93	92	92	88	99	100	100	98	95	100	100	100	100	99
0816	1	71	74	76	60	60	81	84	86	72	79	93	95	97	86	91
0825	1	76	89	86	78	65	84	94	93	87	83	92	95	97	94	90
0803	1	53	54	54	73	60	79	73	73	83	78	86	91	90	95	92
0828	1	42	45	46	48	50	88	87	87	80	80	96	95	95	93	90
0810	2	65	87	59	45	50	78	88	73	63	70	89	90	84	79	85
0822	1	94	98	91	85	80	99	91	99	95	92	100	91	100	98	94
0823	2	67	67	66	74	64	82	82	82	86	84	92	92	91	94	94
0807	1	58	58	57	56	55	78	78	76	73	75	90	90	89	88	88
0801	1	86	70	65	78	65	93	83	78	89	82	100	97	93	96	92
0826	1	68	44	58	64	60	93	72	83	87	80	99	92	94	97	92
0824	1	53	56	58	67	--	72	73	75	84	--	90	91	92	93	--
0805	1	64	52	54	--	--	87	77	77	--	--	96	92	92	--	--
0809	1	54	54	59	69	65	75	75	77	86	80	90	90	90	96	92
Mean		70.1	71.5	70.6	68.9	64.1	87.1	86.0	86.7	85.1	81.3	95.2	93.8	94.9	93.9	91.8
SD		14.0	16.3	14.1	13.0	10.5	9.6	9.4	9.9	9.5	6.4	4.5	3.9	4.5	5.5	3.2
Median		68.8	72.7	71.1	69.2	60.0	88.0	87.5	86.3	86.2	80.0	96.3	91.2	95.8	94.9	92.0

[1] Excludes EST and HOTS Items.

[2] Educator: 1=Yes; 2=No

Table 29. Summary of Judges' Five Sets of Achievement Levels for the Reduced[1] Item Pool (Grade 12, 19 Judges)

		Basic					Proficient					Advanced				
ID	ED[2]	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1213	1	36	36	45	--	--	85	90	91	--	--	98	100	100	--	--
1215	2	39	51	55	--	--	72	82	87	--	--	90	97	99	--	--
1221	2	39	40	39	51	60	78	80	81	74	75	96	96	96	93	90
1208	2	21	23	26	--	--	57	65	69	--	--	96	98	95	--	--
1212	1	46	46	46	51	53	72	73	73	82	80	90	90	90	92	91
1202	2	35	38	39	--	--	80	78	79	--	--	94	94	95	--	--
1223	1	53	52	49	42	58	95	94	86	83	80	100	99	96	92	90
1203	1	76	73	65	--	--	87	84	82	--	--	93	91	91	--	--
1210	1	43	44	45	52	55	77	77	78	78	75	92	92	92	92	90
1219	1	76	74	70	--	--	89	87	85	--	--	96	96	95	--	--
1205	1	58	55	57	40	50	80	79	82	74	75	94	93	91	94	90
1222	1	38	43	43	45	52	68	71	72	71	72	88	89	90	91	92
1204	2	72	67	64	--	--	92	91	91	--	--	98	99	100	--	--
1217	1	51	56	55	56	60	79	81	81	86	80	98	98	98	96	90
1220	2	68	55	63	--	--	89	76	83	--	--	96	93	95	--	--
1209	2	36	37	43	--	--	78	79	79	--	--	93	94	94	--	--
1201	2	79	76	73	73	65	90	87	86	86	85	98	96	95	94	94
1207	1	40	49	54	80	55	71	79	81	93	80	91	94	89	99	90
1206	1	63	61	55	--	--	87	83	79	--	--	96	93	92	--	--
Mean		51.0	51.2	51.9	54.4	56.4	80.3	80.8	81.2	81.0	78.0	94.7	94.8	94.8	93.4	90.8
SD		17.0	14.3	11.9	13.6	4.7	9.6	7.2	5.6	7.2	4.0	3.3	3.1	3.1	2.7	1.4
Median		45.6	50.7	54.2	51.5	55.0	79.9	80.2	81.1	82.4	80.0	95.9	94.0	94.9	92.1	90.0

[1] Excludes EST and HOTS Items.

[2] Educator: 1=Yes; 2=No

Table 30. Summary of Achievement Levels for Content Categories Based Upon (Adjusted)[1] Fourth Round Ratings (Reduced Item Pool)

Grade	Content Category	# of Items	Achievement Levels		
			Basic	Proficient	Advanced
8	Numbers and Operations	52	55%	80%	93%
	Measurement	20	48%	77%	92%
	Geometry	14	41%	66%	82%
	Data Analysis, Statistics, and Probability	9	45%	74%	89%
	Algebra and Functions	14	48%	75%	89%
8	Numbers and Operations	46	64%	82%	93%
	Measurement	21	65%	82%	93%
	Geometry	26	56%	78%	91%
	Data Analysis, Statistics, and Probability	19	58%	79%	91%
	Algebra and Functions	25	54%	78%	91%
12	Numbers and Operations	37	65%	86%	92%
	Measurement	23	58%	82%	91%
	Geometry	24	55%	82%	91%
	Data Analysis, Statistics, and Probability	22	59%	81%	90%
	Algebra and Functions	38	31%	68%	85%

---

[1] Adjusted to be in line with the final recommended achievement levels in the December 18 memo to Roy Truby

[2] Excludes EST and HOTS Items.

Table 31. Summary of Achievement Levels for Mathematics Abilities Based Upon (Adjusted) Fourth Round Ratings (Reduced[1] Item Tool)

			Achievement Levels		
Grade	Process	# of Items	Basic	Proficient	Advanced
4	Conceptual Understanding	40	53.7	79.9	91.4
	Procedural Knowledge	33	54.3	81.6	92.5
	Problem-Solving	36	43.1	73.7	87.8
8	Conceptual Understanding	59	65.2	80.2	90.6
	Procedural Knowledge	41	67.2	82.8	92.0
	Problem-Solving	37	58.1	77.4	88.8
12	Conceptual Understanding	53	60.4	79.8	86.8
	Procedural Knowledge	48	59.5	81.8	91.7
	Problem-Solving	43	46.3	72.0	87.0

[1] Excludes estimation (EST) and higher order thinking skills (HOTS) Items.

Table 32. Analysis of Grade 4 Item Appropriateness Ratings (N=10)

			Item Appropriateness Rating[1]			Statistics	
Content Category	Item	Page	1	2	3	$\bar{x}$	SD
Numbers and Operations	1	1	1	0	9	2.8	.63
	2	2	1	1	8	2.7	.68
	3	3	0	1	9	2.9	.32
	4	5	0	3	7	2.7	.48
	5	6	0	2	8	2.8	.42
	6	7	0	3	7	2.7	.48
	7	9	0	2	8	2.8	.42
	8	11	1	1	8	2.7	.68
	9	13	1	3	6	2.5	.71
	10	14	0	1	9	2.9	.32
	11	15	0	3	7	2.7	.48
	12	16	0	2	8	2.8	.42
	13	17	1	3	6	2.5	.71
	14	18	0	2	8	2.8	.42
	15	19	0	3	7	2.7	.48
	16	20	0	2	8	2.8	.42
	17	21	1	1	8	2.7	.68
	18	22	0	3	7	2.7	.48
	19	23	0	2	8	2.8	.42
	20	24	0	4	6	2.6	.52
	21	25	0	2	8	2.8	.42
	22	26	0	2	8	2.8	.42
	23	28	0	1	9	2.9	.32
	24	30	0	1	9	2.9	.32
	25	32	0	4	6	2.6	.52
	26	33	0	4	6	2.6	.52
	27	34	1	2	7	2.6	.70
	28	36	1	2	7	2.6	.70
	29	38	1	2	7	2.6	.70
	30	39	1	1	8	2.7	.68
	31	41	1	4	5	2.4	.70
	32	42	1	1	8	2.7	.68
	33	43	0	3	7	2.7	.48
	34	44	0	2	8	2.8	.42
	35	45	0	3	7	2.7	.48
	36	46	0	4	6	2.6	.52
	37	48	0	2	8	2.8	.42
	38	49	0	4	6	2.6	.52
	39	50	0	3	7	2.7	.48
	40	52	0	3	7	2.7	.48
	41	53	0	2	8	2.8	.42
	42	54	0	3	7	2.7	.48
	43	55	1	4	5	2.4	.70
	44	56	0	3	7	2.7	.48
	45	57	0	3	7	2.7	.48
	46	58	0	3	7	2.7	.48
	47	59	0	2	8	2.8	.42
	48	60	0	3	7	2.7	.48
	49	61	0	2	8	2.8	.42
	50	62	0	4	6	2.6	.52
	51	63	0	1	9	2.9	.32
	52	64	0	4	6	2.6	.52
	53	65	0	5	5	2.5	.53
	54	66	1	3	6	2.5	.71
	55	67	0	4	6	2.6	.52
	56	68	0	4	6	2.6	.52

Table 32. (Continued)

Analysis of Grade 4 Item Appropriateness Ratings (N=10)

			Item Appropriateness Rating[1]			Statistics	
Content Category	Item	Page	1	2	3	$\bar{x}$	SD
	57	69	0	5	5	2.5	53
	58	70	0	2	8	2.8	42
	59	71	0	1	9	2.9	.32
	60	72	0	3	7	2.7	.48
	61	73	0	3	7	2.7	.48
	62	74	0	4	6	2.6	.52
	63	75	1	3	6	2.5	.71
	64	76	1	2	7	2.6	.70
	65	77	1	2	7	2.6	.70
	66	78	1	3	6	2.5	.71
	67	79	0	3	7	2.7	.48
Measurement	68	80	1	5	4	2.3	.68
	69	81	1	4	5	2.4	.70
	70	82	0	2	8	2.8	.42
	71	83	0	3	7	2.7	.48
	72	84	0	3	7	2.7	.48
	73	85	0	3	7	2.7	.48
	74	86	0	2	8	2.8	.42
	75	87	1	0	9	2.8	.63
	76	89	0	2	8	2.8	.42
	77	90	0	2	8	2.8	.42
	78	91	0	3	7	2.7	.48
	79	93	0	4	6	2.6	.52
	80	95	1	3	6	2.5	.71
	81	96	0	3	7	2.7	.48
	82	97	0	4	6	2.6	.52
	83	98	0	5	5	2.5	.53
	84	100	0	2	8	2.8	.42
	85	102	1	3	6	2.5	.71
	86	103	0	7	3	2.3	.48
	87	104	1	2	7	2.6	.70
	88	105	1	1	8	2.7	.68
	89	106	0	1	9	2.9	.32
	80	107	0	1	9	2.9	.32
	91	108	1	2	7	2.6	.70
	92	110	0	3	7	2.7	.48
	93	111	0	3	7	2.7	.48
	94	112	0	1	9	2.9	.32
	95	113	0	4	6	2.6	.52
	96	115	0	3	7	2.7	.48
Geometry	97	117	0	7	3	2.3	.48
	98	119	3	6	1	1.8	.63
	99	120	0	3	7	2.7	.48
	100	122	1	2	7	2.6	.70
	101	124	2	5	3	2.3	.82
	102	125	0	3	7	2.7	.48
	103	126	0	4	6	2.6	.52
	104	127	0	4	6	2.6	.52
	105	129	0	7	3	2.3	.48
	106	131	0	7	3	2.3	.48
	107	132	0	4	6	2.6	.52
	108	134	3	5	2	1.9	.74
	109	136	2	5	3	2.1	.74
	110	137	0	6	4	2.4	.52
	111	138	0	5	5	2.5	.53
	112	140	1	6	3	2.2	.63

Table 32. (Continued)

Analysis of Grade 4 Item Appropriateness Ratings (N=10)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\bar{x}$	SD
Algebra and Functions	113	142	0	3	7	2.7	.48
	114	143	3	4	3	2.0	.82
	115	144	0	3	7	2.7	.48
	116	145	0	3	7	2.7	.48
	117	146	0	2	8	2.8	.42
	118	147	0	3	7	2.7	.48
	119	148	0	2	8	2.8	.42
	120	150	0	3	7	2.7	.48
	121	151	0	2	8	2.8	.42
	122	152	0	1	9	2.9	.32
	123	153	0	3	7	2.7	.48
	124	155	0	2	8	2.8	.42
	125	156	0	2	8	2.8	.42
	126	157	0	3	7	2.7	.48
	127	159	0	2	8	2.8	.42
	128	161	0	4	6	2.6	.52
	129	163	0	2	8	2.8	.42
	130	165	0	2	8	2.8	.42
	131	167	0	2	8	2.8	.42
	132	168	0	3	7	2.7	.48
	133	170	0	2	8	2.8	.42
	134	171	0	5	5	2.5	.53
	135	172	0	2	8	2.8	.42
	136	174	1	1	8	2.7	.68
	137	176	0	5	5	2.5	.53
	138	177	2	4	4	2.2	.79
	139*	179	1	2	6	2.6	.73
	140	180	0	2	8	2.8	.42
	141	181	0	2	8	2.8	.42
	142	182	0	4	6	2.6	.52
	143	183	1	1	8	2.7	.68

[1] Item Appropriateness Rating: 1 = Low, 2 = Medium, 3 = High

*Only nine judges provided ratings.

Table 33. Analysis of Grade 8 Item Appropriateness Ratings (N=18)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\bar{x}$	SD
Numbers and Operations	1	1	5	7	6	2.1	.80
	2	2	1	6	11	2.6	.62
	3*	3	4	7	6	2.1	.78
	4	5	2	9	7	2.3	.67
	5*	6	0	7	10	2.6	.51
	6	7	2	10	6	2.2	.65
	7	8	5	6	7	2.1	.83
	8	9	7	6	5	1.9	.83
	9	11	8	4	6	1.9	.90
	10	13	3	5	10	2.4	.78
	11	15	1	5	12	2.6	.61
	12	16	2	6	10	2.4	.70
	13	17	0	9	9	2.5	.51
	14	18	1	6	11	2.6	.62
	15	19	6	6	6	2.0	.84
	16	20	4	9	5	2.1	.72
	17	21	1	8	9	2.4	.62
	18**	22	3	5	8	2.3	.79
	19*	24	4	7	6	2.1	.78
	20	25	1	6	11	2.6	.62
	21	26	0	7	11	2.6	.50
	22*	27	0	6	11	2.6	.49
	23	28	0	6	12	2.7	.48
	24	29	2	3	13	2.6	.70
	25	30	3	7	8	2.3	.75
	26	31	9	4	5	1.8	.88
	27	32	8	4	6	1.9	.90
	28	33	3	8	7	2.2	.73
	29*	35	6	4	7	2.1	.90
	30	36	10	3	5	1.7	.90
	31	38	2	9	7	2.3	.67
	32	40	3	9	6	2.2	.71
	33	41	4	4	10	2.3	.84
	34	42	4	9	5	2.1	.72
	35*	43	3	5	9	2.4	.79
	36	44	4	9	5	2.1	.72
	37	45	2	9	7	2.3	.67
	38	46	3	7	8	2.3	.75
	39	47	2	7	9	2.4	.70
	40	48	2	6	10	2.4	.70
	41	49	3	10	5	2.1	.68
	42	51	2	11	5	2.2	.62
	43	53	1	11	6	2.3	.58
	44*	53A	4	6	7	2.2	.81
	45	54	1	9	8	2.4	.61
	46	55	1	8	9	2.4	.62
	47	56	2	9	7	2.3	.67
	48	57	1	11	6	2.3	.58
	49*	58	1	4	12	2.6	.61
	50	59	2	6	10	2.4	.70
	51	60	1	8	9	2.4	.62
	52	61	1	8	9	2.4	.62
	53	62	2	8	8	2.3	.69
	54	63	2	9	7	2.3	.67
	55	64	3	6	9	2.3	.77
	56	65	1	9	8	2.4	.61

Table 33. (Continued)

Analysis of Grade 8 Item Appropriateness Ratings (N=18)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\overline{x}$	SD
	57	66	3	10	5	2.1	.68
	58	68	4	5	9	2.3	.83
	59	69	3	10	5	2.1	.68
	60	70	5	8	5	2.0	.77
	61	71	4	7	7	2.2	.79
	62	72	0	7	11	2.6	.50
	63	73	7	5	6	1.9	.87
	64	74	5	6	7	2.1	.83
	65	75	3	8	7	2.2	.73
	66	76	7	10	1	2.6	.51
	67	77	2	10	6	2.2	.65
	68**	78	1	9	6	2.3	.60
	69	79	1	10	7	2.3	.59
	70	80	4	9	5	2.1	.72
	71	81	2	6	10	2.4	.70
	72	83	3	7	8	2.3	.75
	73	84	2	5	11	2.5	.71
	74*	85	1	8	8	2.4	.62
	75**	86	1	7	8	2.4	.63
	76*	87	0	8	9	2.5	.51
	77**	88	1	8	7	2.4	.62
Measurement	78*	89	0	7	10	2.6	.51
	79*	90	6	6	6	2.0	.84
	80	91	8	5	5	1.8	.86
	81	92	7	6	5	1.9	.83
	82	93	2	8	8	2.3	.69
	83	94	1	6	11	2.6	.62
	84*	95	3	6	8	2.3	.77
	85	96	5	9	4	1.9	.72
	86	97	1	8	9	2.4	.62
	87	98	1	6	11	2.6	.62
	88	99	3	10	5	2.1	.68
	89	100	5	8	5	2.0	.77
	90*	102	3	7	7	2.2	.75
	91	104	5	8	5	2.0	.77
	92	105	5	8	5	2.0	.77
	93*	107	3	9	5	2.1	.70
	94	108	2	7	9	2.4	.70
	95*	109	2	8	7	2.3	.69
	96	110	6	5	7	2.1	.87
	97	111	3	11	4	2.1	.64
	98	112	1	7	10	2.5	.62
	99	113	7	4	7	2.0	.91
	100	114	6	6	6	2.0	.84
	101*	115	2	4	11	2.5	.72
	102*	116	0	9	8	2.5	.51
	103	117	2	5	11	2.5	.71
	104	118	3	7	8	2.3	.75
	105	119	2	8	8	2.3	.69
	106*	121	3	7	7	2.2	.75
	107	122	4	8	6	2.1	.76
	108*	123	2	6	9	2.4	.71
	109	124	3	6	9	2.3	.77
	110	125	0	8	10	2.6	.51
	111*	126	0	5	12	2.7	.47
Geometry	112	128	0	5	13	2.7	.46

Table 33. (Continued)

Analysis of Grade Item Appropriateness Ratings (N=18)

			Item Appropriateness Rating[1]			Statistics	
Content Category	Item	Page	1	2	3	$\bar{x}$	SD
	113	129	6	7	5	1.9	.80
	114	130	7	8	3	1.8	.73
	115	131	5	8	5	2.0	.77
	116*	132	2	7	8	2.4	.70
	117	134	4	10	4	2.0	.69
	118	136	4	6	8	2.2	.81
	119*	137	2	9	6	2.2	.66
	120	138	4	6	8	2.2	.81
	121*	139	6	5	5	1.9	.85
	122	140	4	7	7	2.2	.79
	123	141	4	8	6	2.1	.76
	124	142	1	6	11	2.6	.62
	125	143	2	7	9	2.4	.70
	126*	144	1	9	7	2.4	.61
	127	145	2	10	6	2.2	.65
	128	147	4	3	11	2.4	.85
	129	149	1	9	8	2.4	.61
	130*	150	3	6	8	2.3	.77
	131	151	1	7	10	2.5	.62
	132	152	0	7	11	2.6	.50
	133	154	3	7	8	2.3	.75
	134**	156	0	7	9	2.6	.51
	135	158	1	4	13	2.7	.59
	136	160	1	4	13	2.7	.59
	137	161	2	8	8	2.3	.69
	138	162	0	8	10	2.6	.51
	139	163	1	5	12	2.6	.61
	140*	164	1	7	9	2.5	.62
Data Analysis, Statistics, and Probability	141	166	0	9	9	2.5	.51
	142	167	6	7	5	1.9	.80
	143	168	4	5	9	2.3	.83
	144	170	3	5	10	2.4	.78
	145	172	2	6	10	2.4	.70
	146	173	2	5	11	2.5	.71
	147	174	5	3	10	2.3	.90
	148	175	4	9	5	2.1	.72
	149	176	0	8	10	2.6	.51
	150*	178	7	5	5	1.9	.86
	151	179	4	10	4	2.0	.69
	152	180	5	6	7	2.1	.83
	153	181	2	10	6	2.2	.65
	154	183	1	8	9	2.4	.62
	155	184	1	6	11	2.6	.62
	156	185	0	9	9	2.5	.51
	157	186	2	7	9	2.4	.70
	158	187	4	9	5	2.1	.72
	159	188	0	8	10	2.6	.51
	160	189	3	5	10	2.4	.78
	161	191	0	6	12	2.7	.48
	162*	193	2	5	10	2.5	.72
	163	194	1	11	6	2.3	.58
	164	196	3	6	9	2.3	.77
Algebra and Functions	165	198	6	11	1	2.6	.49
	166	199	2	8	8	2.3	.69
	167	200	1	7	10	2.5	.62
	168	201	4	5	9	2.3	.83

Table 33. (Continued)

Analysis of Grade 8 Item Appropriateness Ratings (N=18)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\bar{x}$	SD
	169	202	1	10	7	2.3	.59
	170	203	6	7	5	1.9	.80
	171*	205	1	8	8	2.4	.62
	172*	207	3	8	6	2.2	.73
	173*	208	0	6	11	2.6	.49
	174*	209	1	4	12	2.6	.61
	175	211	4	5	9	2.3	.83
	176*	212	1	5	11	2.6	.62
	177	213	6	6	6	2.0	.84
	178	214	4	6	8	2.2	.81
	179	216	2	4	12	2.6	.70
	180	217	2	7	9	2.4	.70
	181	218	7	6	5	1.9	.83
	182	219	2	8	8	2.3	.69
	183	220	7	5	6	1.9	.87
	184**	222	3	6	7	2.2	.78
	185	224	4	8	6	2.1	.76
	186*	226	5	7	5	2.0	.79
	187	227	4	10	4	2.0	.69
	188	228	6	7	5	1.9	.80
	189	229	1	4	13	2.7	.59
	190	230	0	6	12	2.7	.48
	191*	231	2	5	10	2.5	.72

[1] Item Appropriateness Rating: 1 = Low, 2 = Medium, 3 = High

*Only 17 judges provided ratings.

**Only 16 judges provided ratings.

Table 34. Analysis of Grade 12 Item Appropriateness Ratings (N=8)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\overline{x}$	SD
Numbers and Options	1	1	1	2	5	2.5	.76
	2	2	1	2	5	2.5	.76
	3	3	0	5	3	2.4	.52
	4	4	1	1	6	2.6	.74
	5	5	1	1	6	2.6	.74
	6*	6	1	4	2	2.8	.43
	7	8	2	1	5	2.4	.92
	8	10	0	2	6	2.6	.43
	9	12	1	2	5	2.5	.76
	10	13	2	2	4	2.2	.89
	11	14	0	1	7	2.9	.35
	12	15	1	4	3	2.2	.71
	13	16	1	3	4	2.4	.74
	14	17	0	4	4	2.5	.54
	15	18	0	4	4	2.5	.54
	16	19	0	3	5	2.6	.52
	17	20	0	4	4	2.5	.54
	18	21	3	1	4	2.1	.99
	19	22	4	0	4	2.0	1.07
	20	23	3	2	3	2.0	.93
	21	24	4	1	3	1.9	.99
	22	25	0	3	5	2.6	.52
	23	26	1	2	5	2.5	.76
	24	28	0	4	4	2.5	.54
	25	30	0	3	5	2.6	.52
	26	31	5	1	2	1.6	.92
	27	32	0	3	5	2.6	.52
	28	33	2	3	3	2.1	.84
	29	34	1	2	5	2.5	.76
	30	36	0	3	5	2.6	.52
	31	37	1	3	4	2.4	.74
	32	38	0	3	5	2.6	.52
	33	39	0	2	6	2.8	.46
	34	40	3	1	4	2.1	.99
	35	41	0	2	6	2.8	.46
	36	42	0	2	6	2.8	.46
	37	43	0	3	5	2.6	.52
	38	44	2	1	5	2.4	.92
	39	45	0	1	7	2.9	.35
	40	46	1	3	4	2.4	.74
	41	47	0	3	5	2.6	.52
	42	48	1	3	4	2.4	.74
	43	49	0	3	5	2.6	.52
	44	50	1	1	6	2.6	.74
	45	51	0	2	6	2.8	.46
	46	52	0	3	5	2.6	.52
	47	53	0	3	5	2.6	.52
	48	54	1	3	4	2.4	.74
	49	55	1	2	5	2.5	.76
	50	57	2	3	3	2.1	.84
	51	59	1	2	5	2.5	.76
	52	60	2	1	5	2.4	.92
	53	61	0	4	4	2.5	.54
	54	62	3	1	4	2.1	.99
	55	63	1	2	5	2.5	.76
	56	65	1	2	5	2.5	.76
	57	66	1	0	7	2.8	.71

Table 34. (Continued)

Analysis of Grade 12 Item Appropriateness Ratings (N=18)

Content Category	Item	Page	Item Appropriateness Rating[1]			Statistics	
			1	2	3	$\overline{X}$	SD
Measurement	58	67	0	5	3	2.4	.52
	59	68	0	1	7	2.9	.35
	60	69	0	2	6	2.8	.46
	61	70	0	2	6	2.8	.46
	62	71	0	3	5	2.6	.52
	63	72	3	2	3	2.0	.93
	64	73	1	3	4	2.4	.74
	65	74	1	1	6	2.6	.74
	66	75	0	2	6	2.8	.46
	67	76	2	2	4	2.2	.89
	68	77	5	0	3	1.8	1.04
	69	78	2	1	5	2.4	.92
	70	79	0	4	4	2.5	.54
	71	80	3	3	2	1.9	.84
	72	81	0	2	6	2.8	.46
	73	82	0	2	6	2.8	.46
	74	83	3	2	3	2.0	.93
	75	84	1	4	3	2.2	.71
	76	85	1	1	6	2.6	.74
	77	87	0	2	6	2.8	.46
	78	88	2	0	6	2.5	.93
	79	89	1	3	4	2.4	.74
	80	90	1	2	5	2.5	.76
	81	91	1	2	5	2.5	.76
	82	93	1	3	4	2.4	.74
	83	94	3	2	3	2.0	.93
	84	95	2	1	5	2.4	.92
	85	96	2	5	1	1.9	.64
	86	97	2	5	1	1.9	.64
	87	98	1	2	5	2.5	.76
	88	99	1	2	5	2.5	.76
	89	100	0	4	4	2.5	.54
	90	101	1	3	4	2.4	.74
	91	102	2	2	4	2.2	.89
	92	103	0	3	5	2.6	.52
	93	104	0	2	6	2.8	.46
	94	106	0	3	5	2.6	.52
	95	108	0	3	5	2.6	.52
	96	109	0	2	6	2.8	.46
Geometry	97*	110	2	1	4	2.3	.95
	98	112	0	2	5	2.7	.49
	99	113	0	1	6	2.9	.38
	100*	115	0	1	6	2.9	.38
	101*	116	0	1	6	2.9	.38
	102*	117	1	2	4	2.4	.79
	103*	118	1	1	5	2.6	.79
	104*	119	0	2	5	2.7	.49
	105*	120	2	2	3	2.1	.90
	106*	121	2	2	3	2.1	.90
	107*	122	0	2	5	2.7	.49
	108*	123	0	4	3	2.4	.54
	109*	124	0	3	4	2.6	.54
	110*	126	0	3	4	2.6	.54
	111*	127	0	5	2	2.3	.49
	112*	128	2	2	3	2.1	.90
	113	129	1	3	4	2.4	.74

Table 34. (Continued)

Analysis of Grade 12 Item Appropriateness Ratings (N=8)

			Item Appropriateness Rating[1]			Statistics	
Content Category	Item	Page	1	2	3	$\overline{x}$	SD
	114	130	3	2	3	2.0	.93
	115	131	2	3	3	2.1	.84
	116	132	1	4	3	2.2	.71
	117	133	1	3	4	2.4	.74
	118	135	0	3	5	2.6	.52
	119	137	0	3	5	2.6	.52
	120	138	1	3	4	2.4	.74
	121	139	0	3	5	2.6	.52
	122	140	2	1	5	2.4	.92
	123	141	1	3	4	2.4	.74
	124	142	1	3	4	2.4	.74
	125	143	0	2	6	2.8	.46
	126	144	0	3	5	2.6	.52
	127	146	1	4	3	2.2	.71
	128	147	0	4	4	2.5	.54
	129	149	0	1	7	2.9	.35
	130	151	1	2	5	2.5	.76
	131	152	3	1	4	2.1	.99
	132	153	0	3	5	2.6	.52
Data Analysis, Statistics, and Probability	133	155	2	1	5	2.4	.92
	134	156	1	1	6	2.6	.74
	135	158	0	2	6	2.8	.46
	136	160	1	1	6	2.6	.74
	137	162	0	4	4	2.5	.54
	138	164	1	2	5	2.5	.76
	139	165	1	3	4	2.4	.74
	140	166	0	2	6	2.8	.46
	141	167	0	2	6	2.8	.46
	142	168	1	3	4	2.4	.74
	143	169	1	2	5	2.5	.76
	144	170	2	3	3	2.1	.84
	145	172	1	1	6	2.6	.74
	146	173	0	3	5	2.6	.52
	147	174	1	2	5	2.5	.76
	148	176	0	1	7	2.9	.35
	149	177	1	2	5	2.5	.76
	150	178	2	1	5	2.4	.92
	151	179	0	4	4	2.5	.54
	152	181	0	2	6	2.8	.46
	153	182	0	3	5	2.6	.52
	154	183	0	3	5	2.6	.52
	155	185	0	4	4	2.5	.54
	156	186	2	2	4	2.2	.89
	157	188	1	2	5	2.5	.76
	158	189	1	1	6	2.6	.74
	159	191	0	2	6	2.8	.46
Algebra and Functions	160	193	0	1	7	2.9	.35
	161	194	0	3	5	2.6	.52
	162	195	2	1	5	2.4	.92
	163	196	0	5	3	2.4	.52
	164	197	0	5	3	2.4	.52
	165	198	1	3	4	2.4	.74
	166	199	1	2	5	2.5	.76
	167	200	0	3	5	2.6	.52
	168	202	0	2	6	2.8	.46
	169	203	0	3	5	2.6	.52
	170	205	1	2	5	2.5	.76

Table 34. (Continued)

Analysis of Grade 12 Item Appropriateness Ratings (N=8)

Content Category	Item	Page	Item Appropriateness Rating[1] 1	2	3	Statistics $\overline{x}$	SD
	171	206	0	2	6	2.8	.46
	172	207	3	1	4	2.1	.99
	173	208	0	2	6	2.8	.46
	174	209	0	3	5	2.6	.52
	175	210	1	3	4	2.4	.74
	176	211	2	2	4	2.2	.89
	177	213	0	3	5	2.6	.52
	178	214	0	4	4	2.5	.54
	179	215	1	4	3	2.2	.71
	180	216	1	4	3	2.2	.71
	181	217	1	3	4	2.4	.74
	182	218	0	4	4	2.5	.54
	183	220	0	4	4	2.5	.54
	184	221	2	4	2	2.0	.76
	185	222	1	3	4	2.4	.74
	186	224	1	2	5	2.5	.76
	187	225	0	3	5	2.6	.52
	188	227	1	2	5	2.5	.76
	189	229	1	5	2	2.1	.64
	190	231	1	2	5	2.5	.76
	191	233	2	2	4	2.2	.89
	192	235	2	2	4	2.2	.89
	193	237	2	1	5	2.4	.92
	194	238	0	4	4	2.5	.54
	195	240	0	3	5	2.6	.52
	196	242	0	2	6	2.8	.46
	197	244	1	2	5	2.5	.76
	198	245	0	4	4	2.5	.54
	199	247	1	3	4	2.4	.74
	200	248	1	3	4	2.4	.74
	201	249	0	3	5	2.6	.52
	202	250	0	4	4	2.5	.54
	203	251	1	3	4	2.4	.74

[1] Item Appropriateness Rating: 1 = Low, 2 = Medium, 3 = High

*Only seven judges provided ratings.

Table 35. Summary of Mean Item Appropriateness Ratings

Grade	Number of Items	Number of Judges	Distribution of Means Low (1.00-1.49)	Medium (1.50-2.49)	High (2.50-3.00)
4	143	10	0%	11.2%	88.8%
8	191	18	0%	73.3%	26.7%
12	203	8	0%	37.4%	62.6%

Table 36. Correlations Between First, Second, and Third Round of Average Judges' Ratings of Expected Item p-Values and Actual p-Values

		Correlation		
Grade	Level	r_{1p}	r_{2p}	r_{3p}
4	Basic	.26	.46	.45
	Proficient	.26	.48	.46
	Advanced	.23	.47	.45
8	Basic	.63	.76	.77
	Proficient	.60	.77	.76
	Advanced	.57	.76	.72
12	Basic	.78	.88	.88
	Proficient	.79	.89	.87
	Advanced	.75	.81	.78

Table 37. Summary of Grade 4 First Round Achievement Levels, Reported for Groups (N=22)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1	1st	40.0	21.6	62.2	20.3	78.2	16.2
2	1st	54.2	13.0	78.5	10.3	92.2	8.1
3	1st	36.4	18.6	62.2	17.0	78.8	15.1
4	1st	62.3	8.2	83.5	4.8	94.5	2.5
T	1st	49	18	72	16	87	13

Table 38. Summary of Grade 8 First Round Achievement Levels Reported for Groups (N=22)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1	1st	81.8	9.2	95.2	6.2	98.0	3.2
2	1st	77.4	9.4	91.6	7.4	97.4	2.9
3	1st	57.0	9.0	79.3	4.2	91.0	2.3
4	1st	64.0	13.1	82.2	11.0	94.2	5.5
T	1st	70	14	87	10	95	4

Table 39. Summary of Grade 12 First Round Achievement Levels Reported for Groups (N=19)

		Achievement Level					
		Basic		Proficient		Advanced	
Group	Item Ratings	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1	1st	67.0	17.7	85.0	9.4	95.5	3.1
2	1st	49.8	12.5	83.8	5.5	95.2	3.4
3	1st	46.4	18.4	74.8	10.0	93.4	3.8
4	1st	52.8	13.2	82.4	6.8	95.8	3.8
T	1st	53	16	81	9	95	3.0

Appendix G

Technical Memo

UNIVERSITY OF MASSACHUSETTS
AT AMHERST

Hills House
Amherst, MA 01003
(413) 545-0262

Laboratory of Psychometric and
Evaluative Research

FROM: Ronald K. Hambleton RKH
University of Massachusetts at
Amherst

DATE: December 18, 1990

TO: Roy Truby, Executive Director

CONCERNING: Recommended Adjustments in the Grades 4, 8, and 12
Achievement Levels

In my haste to complete the December 13th memo for the meeting on December 17th, a number of minor errors in my calculations went undetected. In addition, a number of other points were not made as clearly or accurately as they should have been. Please substitute this edited memo for the one I mailed to you a few days ago. Also, I have added a postscript to this memo that summarizes the views of the Technical Advisory Committee on Standard Setting (TACSS). The committee and I were in complete agreement on the postscript. I should add that Dick Jaeger was unable to be present at the meeting yesterday and so his views are unknown at this time.

When our TACSS met in Washington on October 30, 1990, we reviewed the statistical data that were available at the time and discussed a number of problems including a few skewed distributions of achievement levels (notably at the grade 8 level) and the inappropriate inclusion of EST and HOTS items in the calculation of achievement levels. At that time, the TACSS felt that I should consider:

(1) adjustments necessitated by the separate reporting of performance on the EST and HOTS items from other items in the item pool;

(2) substitution of the median ratings for the mean ratings to more adequately reflect central tendency with skewed distributions of judges' ratings;

(3) adjustments due to the non-participation of 40% of the judges at the Washington meeting;

(4) "smoothing" of the achievement levels on the NAEP reporting scale due to (possible) inconsistencies.

In this report, I will describe my recommended adjustments based upon a consideration of the first three points above. The fourth point can be considered in more detail once we have the mappings of achievement levels in the five content areas onto the NAEP reporting scale at each grade level and some details from ETS on the method of aggregation of content scores

The University of Massachusetts is an Affirmative Action/Equal Opportunity Institution

into composite scores. (Let me add that Gene Johnson provided the information I need at our meeting yesterday.)

Adjustments Due to Deletion of the EST and HOTS Items

The achievement levels based upon the total and reduced NAEP item pool are presented in Tables 1 to 3 and 1A to 3A. Making adjustments due to the deletion of EST and HOTS items from the item pool is complicated by one factor: item ratings were not available on the fifth and final round (recall that judges provided overall ratings at round 5), therefore achievement levels could not be calculated directly for the reduced item pool.

The solution I came up with was to calculate the differences between achievement levels on the fourth round for the total and reduced item pools. Then, I assumed that similar differences would have existed on the fifth and final round, if such differences had been possible to compute. Accordingly, I revised the fifth and final ratings to reflect these differences. These calculations, which were based on statistics in Tables 1 to 3 and 1A to 3A, are shown below:

		Round 4			Round 5	
Grade	Level	Total Pool	Reduced Pool	Difference	Total Pool	Reduced Pool*
4	Basic	48.5	49.4	+0.9	50.3	51.2
	Proficient	76.0	76.5	+0.5	77.3	77.8
	Advanced	89.2	89.6	+0.4	90.2	90.6
8	Basic	68.5	68.9	+0.4	64.1	64.5
	Proficient	84.9	85.1	+0.2	81.3	81.5
	Advanced	93.8	93.9	+0.1	91.8	91.9
12	Basic	56.4	54.4	-2.0	56.4	54.4
	Proficient	82.5	81.0	-1.5	78.0	76.5
	Advanced	93.8	93.4	-0.4	90.8	90.4

*Adjusted for the small differences noted at Round 4 between achievement levels on the total and reduced item pools.

Note that the differences on the round 4 data were small and ranged from -2.0% (grade 12, Basic) to 0.9% (grade 4, Basic).

I wondered whether the differences (adjustments) would look any different if they were based on the third round of ratings. The third round of ratings were obtained from the total group of judges and there was no reason I could think of to expect that the size of the difference between the achievement levels for the total and reduced item pools would be affected by the round at which the differences were estimated. The round 3 differences are shown below:

		Round 3			Round 5	
Grade	Level	Total Pool	Reduced Pool	Difference	Total Pool	Reduced Pool
4	Basic	47.2	47.9	+0.7	50.3	51.0
	Proficient	71.9	72.7	+0.8	77.3	78.1
	Advanced	88.1	88.8	+0.7	90.2	90.9
8	Basic	70.9	70.6	-0.3	64.1	63.8
	Proficient	86.6	86.7	+0.1	81.3	81.4
	Advanced	95.0	94.9	-0.1	91.8	91.7
12	Basic	54.2	51.9	-2.3	56.4	54.1
	Proficient	82.2	81.2	-1.0	78.0	77.0
	Advanced	95.1	94.8	-0.3	90.8	90.5

The adjustments, based on round 3, were very close to those based on the round 4 ratings. Note, too, that the largest difference (grade 12, Basic) did hold up on the Round 3 data.

One adjustment possibility seemed reasonable based upon the analyses above:

1. Make adjustments which are based on the average of the differences at Rounds 3 and 4 between the achievement levels of the total and reduced item pool. These latter adjustments are shown below:

Grade	Level	Fifth Round	Average Difference	Adjusted Levels
4	Basic	50.3	0.8	51.1
	Proficient	77.3	0.7	78.0
	Advanced	90.2	0.6	90.8
8	Basic	64.1	0.1	64.2
	Proficient	81.3	0.2	81.5
	Advanced	91.8	0.0	91.8
12	Basic	56.4	-2.2	54.2
	Proficient	78.0	-1.3	76.7
	Advanced	90.8	-0.4	90.4

In only one instance did the adjustments (after rounding off) move the achievement levels reported in Table 15 by more than 1%. Note, too, that four of the changes are moving achievement levels up by 1% and three of the changes are moving achievement levels down by 1%. I recommend that the above adjustments be made.

Adjustments Due to Skewed Distributions

In Tables 24 to 26, the achievement levels (adjusted for HOTS and EST items in rounds 1 to 4) are reported for all five rounds along with descriptive statistics. A comparison of means and medians highlights the fact that several of the distributions of judges' ratings were skewed (most often, positively skewed), and therefore the median would be a more suitable indicator of central tendency than the mean. While, in standard-setting practice, means are more common than medians, there are important exceptions (e.g., on the NTE exams, see Busch & Jaeger, JEM, 1990). Also, this preference for means in the measurement literature may be due to the presence of homogeneous distributions of judges' ratings. Other possibilities are that standard setters don't look closely at their distributions or give much thought to the matter of means versus medians. In any case, we did look at the means and medians and the statistics are reported below:

Grade	Level	Judges	Mean	Median	Difference
4	Basic	11	50.3	50.0	-0.3
	Proficient	11	77.3	75.0	-2.3
	Advanced	11	90.2	90.0	-0.2
8	Basic	18	64.1	60.0	-4.1
	Proficient	18	81.3	80.0	-1.3
	Advanced	18	91.8	92.0	+0.2
12	Basic	9	56.4	55.0	-1.4
	Proficient	9	78.0	80.0	+2.0
	Advanced	9	90.8	90.0	-0.8

In four of the nine comparisons, the differences were less than 1%; in the other five comparisons, the differences would influence the resulting achievement levels by anywhere from 1% (grade 8, Proficient) to 4% (grade 8, Basic). Four of the adjustments would lower achievement levels and one adjustment (grade 12, Proficient) would raise the achievement levels. I considered looking at the round 4 ratings to see if the trends in the means and medians were the same, but I rejected the idea because of the substantial changes that took place in the distributions of the ratings at rounds 4 and 5. Though the mean or median ratings did not change substantially, the standard deviations did. This was especially true at grades 4 and 12 and therefore making any adjustments in achievement levels due to the skewness of the distributions seemed best left to careful consideration of the fifth and final round of ratings.

I went to Tables 24 to 26 to determine the reasons for the mean versus median differences in the five cases where the difference exceeded 1%.

1. Grade 4 - Proficient (mean = 77.3; median = 75.0). Here the difference was due to two judges whose ratings were about 10% higher than the remainder of the group.

2. Grade 8 - Basic (mean - 64.1; median - 60.0). Here, basically, three judges were 15% to 25% higher than the rest of their group. Interestingly, one of the judges was consistently high; another judge progressively changed in a way opposite to the general trend in the data; and a third judge started with high ratings and then lowered them.

3. Grade 8 - Proficient (mean - 81.3; median - 80.0). The same three judges were also responsible for the positively skewed distributions here, though, because of their very high ratings for the Basic category, there was little room left for them to reflect higher ratings than their fellow judges. As a result, the mean vs. median difference was substantially smaller.

4. Grade 12 - Basic (mean - 56.4; median - 55.0). This small difference appeared to be due to one judge who was above the group average by about 10%.

5. Grade 12 - Proficient (mean - 78.0; median - 80.0). The small difference here seemed to be due to a number of judges providing ratings 5% to 8% below the group average.

The evidence for substituting medians for means in the reporting of achievement levels seems compelling. Four of nine achievement level distributions showed a marked tendency for a small number of judges to be substantially higher in their ratings than other judges, and thereby these judges rendered the mean achievement levels less useful in characterizing the views of the total group of judges. In order that the resulting achievement levels be more representative of the total group of judges, I recommend that the median ratings be substituted for the to mean ratings.

Adjustments Due to Missing Judges in Washington

Tables 18 to 20 provide the relevant information. Thirty-eight of the 63 judges (60%) were present in Washington. But, 25 judges were not present, and the missing judges were mainly the non-educators (12 of 18 did not return to Washington). Other trends in the data (see Table 20) are also clear: The missing judges tended to set somewhat higher standards. Because nearly all of the eighth grade judges returned (19 of 22), I'm suggesting that the grade 8 achievement levels be left as they are. The remainder of the discussion will focus on the grades 4 and 12 results.

The first thing I decided to do was to recalculate the results in Table 20 using the reduced item pools. Table 20 was based on the total item pool, and excluded one judge who provided final ratings late. (I think this judge had to leave the meeting early.) Since the reduced item pool was the appropriate one (see the first section of this memo), I wanted to revise Table 20 to reflect this point. Changes to Table 20 are shown below:

		Not Present in Washington			Present in Washington		
Grade	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD
4	Basic	11	51.8	12.5	11	43.9	8.8
	Proficient	11	76.1	10.1	11	69.4	5.7
	Advanced	11	90.2	5.7	11	87.4	3.8
12	Basic	10	52.5	13.2	9	51.2	9.5
	Proficient	10	82.5	6.6	9	80.0	4.7
	Advanced	10	95.6	3.0	9	93.0	3.1

The means and standard deviations of achievement levels at round 3 are based on the reduced item pools.

The breakdown of educators and non-educators returning to the Washington meeting was as follows:

Grade	Educator	Non-Educator	Not Present in Washington	Present in Washington
4	15		6	9
		7	5	2
12	11		4	7
		8	6	2

Two findings are clear from the results above: (1) Judges not present in Washington tended to set higher standards (especially at grade 4), and (2) two-thirds of the non-educators (or 11 of 15) did not attend the Washington meeting whereas two-thirds of the educators (16 of 26) did attend.

At this point, a number of questions seemed appropriate to ask:

1. Are the differences in achievement levels between the Washington and non-Washington groups on round 3 statistically significant?

   Answer: I suppose that the most powerful method would be a multivariate test of significance, but I was not prepared to invest the time in conducting such an analysis. Instead, I substituted three t-tests (which were not independent and where I used standard deviations obtained by dividing the numerator by N instead of the more correct N-1) at each grade level. The three t-test statistics at grade 4 were 1.72, 1.91, and 1.33 for Basic, Proficient, and Advanced, respectively, which bordered on being statistically significant differences at the .05 level. My guess is that, had I done the analyses with a more powerful statistical method, the observed differences between the two groups would have been found to be statistically significant. At grade 12, the t-statistics were less than 1, except for the Advanced level, where the t-statistic was 1.43. But, in any case, the differences between the groups at grade 4 appeared sizable and in need of some attention.

I decided, therefore, to focus my attention solely on the grade 4 results.

Next, since we lost 5 of the 7 non-educators at grade 4, I wanted to see how achievement levels for educators and non-educators compared at round 3. The questions were:

2. Did educators and non-educators at grade 4 set different achievement levels on the round 3 data?

   and the companion questions:

3. Did the educators who went on to Washington differ from educators who did not? Did the non-educators who went on to Washington differ from non-educators who did not?

The statistical data are shown below:

- Grade 4 Round 3 Data (see Table 24) -

Level	Educator (N=15)	Non-Educator (N=7)	Total (N=22)
Basic	45.9	52.0	47.9
Proficient	70.9	76.6	72.7
Advanced	87.4	91.7	88.8

EDUCATORS

Level	Not Present in Washington (N=6)	Present in Washington (N=9)
Basic	48.3	44.3
Proficient	73.5	69.2
Advanced	88.0	87.0

NON-EDUCATORS

Level	Not Present in Washington (N=5)	Present in Washington (N=2)
Basic	56.0	42.0
Proficient	79.2	70.0
Advanced	92.8	89.0

Of course, the samples are very small but a number of trends in the data are clear:

1. (Question 2). The non-educators set their achievement levels 4 to 6% higher than the educators.

2. (Question 3). Both educators and non-educators who were not present in Washington tended to set higher achievement levels than those who were present in Washington.

Clearly, then, in grade 4, one could speculate that the grade 4 achievement levels would have been higher had the 11 judges who missed the Washington meeting been present. But, it seems possible, too, that these non-educator judges would have been persuaded by other judges that their achievement levels were out-of-line. A cursory look at the six non-educators' ratings who were present in Washington (see Tables 24 to 26) suggested that these six persons tended to revise their ratings in ways that reflected the overall group changes, and therefore the hypothesis is plausible. (The six judges who completed all five rounds of ratings were an average of 6.1% away from the group means on round 3 and 3.8% away from the group means on round 5.) There is also the possibility that, with the missing judges present in Washington, different dynamics may have been set up and the results could have been different.

It is interesting to observe the trends in the data where judges completed the last three rounds of ratings (see above and Table 24):

Grade 4 (N=11)	Round 3	Round 4	Round 5*
Basic	43.9	49.4	51.2
Proficient	69.4	76.5	77.8
Advanced	87.4	89.6	90.6

*Adjusted for the small differences noted at round 4 between achievement levels on the total and reduced item pools.

The judges who were present in Washington did increase their ratings. Was it because they perceived that their ratings were a little low (recall that judges, or many of them, knew the achievement levels from round 3), or did they increase their achievement levels because of some other reasons? The increase was not due to group discussions because these took place between rounds 4 and 5, where the judges showed only small mean changes (always less than 3%) compared to the changes between rounds 3 and 4.

I considered recalculating the fifth round results by using weights to reflect the educator and non-educator balance in Vermont, but the number of available non-educators (2) seemed too small to lead to meaningful results.

What then should be done? One possible recommendation is that there is no defensible way to make the adjustments and, in addition, there is no need to make adjustments. Defensibility for any adjustments is not possible because it is simply impossible to build a psychological model that might explain the impact of the missing judges, and any statistical models which seem reasonable would need to be applied with a small amount of data. Note, too, that the Washington group received additional training, helped to clarify definitions, and spent considerable time discussing their achievement levels with colleagues. Rather than try to defend adjustments, it seems appropriate to defend the Washington meeting and the results that came from it. Eleven judges is a marginally acceptable number of judges and the group is only one short of the desired 30%/70% split of non-educators/educators.

Of course, there is an opposite recommendation that is plausible, too. This recommendation is based on the assumption that a statistical correction is justified because the two grade 4 groups (participants and non-participants in Washington) did differ substantially in their ratings. Any correction is likely to be an overcorrection since the evidence suggests that the judges at each grade level tended to reach a kind of consensus. If it is felt that some adjustments be made, my specific recommendation is that the statistics on page 6 for achievement levels of those present and not present in Washington be used to adjust the final ratings:

Basic: 3.9% (51.8% - 47.9%)
Proficient: 3.4% (76.1% - 72.7%)
Advanced: 1.4% (90.2% - 88.8%)

Other more complicated adjustments could be proposed, but the adjustments on page 6 are straightforward and don't give undue importance to the educator/non-educator distinction, which is only one of several important demographic variables.

Summary

Based upon my analyses of the first three issues (see page 1 of this memo), I believe it is reasonable to recommend adjustments to reflect (1) the reduced item pool and the skewed distributions of judges' ratings at all three grade levels, and (2) the changes in the demographic composition of judges at grade 4 at the Vermont and Washington meetings. The adjustments and recommended achievement levels are shown below:

			- Adjustments -			
Grade	Level	Round 5 Unadjusted	Reduced Item Pool (1)	Substitution of medians (2)	Changing Population (3)	Round 5 Adjusted
4	Basic	50.3	0.8	-0.3	+3.9	54.7
	Proficient	77.3	0.7	-2.3	+3.4	79.1
	Advanced	90.2	0.6	-0.2	+1.4	92.0
8	Basic	64.1	0.1	-4.1	--	60.1
	Proficient	81.3	0.2	-1.3	--	80.2
	Advanced	91.8	0.0	+0.2	--	92.0
12	Basic	56.4	-2.2	-1.4	--	52.8
	Proficient	78.0	-1.3	+2.0	--	78.7
	Advanced	90.8	-0.4	-0.8	--	89.6

I feel comfortable with nearly all of the recommended revisions. The exception is at grade 4 and the proposed adjustments due to changes in the pool of judges between Vermont and Washington. Here, I think a case for other recommendations could be made. I am looking forward to the meeting with the TACSS, ETS staff, and some of your staff to discuss this memo in detail. Perhaps, too, ETS will have prepared the charts I requested for mapping achievement levels onto the NAEP reporting scale. With these charts, we can look at the need for smoothing the data to achieve consistency and coherence across grade levels. One question I want the committee to consider at the meeting concerns standard errors associated with achievement levels. Is there an acceptable way to revise the errors from those reported in Table 16 to reflect the adjustments that are being proposed?

P.S. At our meeting yesterday in Washington, we had an excellent discussion of the points in my memo. Professors Forsyth, Haertel, and myself are in essential agreement about the points in the memo concerning adjustments to the achievement levels due to the removal of HOTS and EST items, and substituting achievement levels based upon median ratings rather than mean ratings. We are also in agreement, after a lengthy discussion, that adjustments should not be made for persons who were unable to be present in Washington to complete the fourth and fifth rounds of ratings. We feel that, at the Washington meeting, definitions were clarified, a revised item rating task was implemented, and valuable and extensive discussions took place among the judges. There is simply no defensible way to predict how judges might have responded had they been present, or the influence they may have had on the ratings of other judges who were present. In addition, the number of judges who were present was at least minimally acceptable and the balance of educators and non-educators at each grade level was at least reasonably close to the desired 30%/70% split. The final recommended achievement levels are given below:

Grade	Level	Round 5 Unadjusted	Adjustment	Round 5 Adjusted	Round 5 Rounded
4	Basic	50.3	0.5	50.8	51
	Proficient	77.3	-1.6	75.7	76
	Advanced	90.2	0.4	90.6	91
8	Basic	64.1	-4.0	60.1	60
	Proficient	81.3	-1.1	80.2	80
	Advanced	91.8	0.2	92.0	92
12	Basic	56.4	-3.6	52.8	53
	Proficient	78.0	0.7	78.7	79
	Advanced	90.8	-1.2	89.6	90

cc.: TACSS Committee Members, Dan Stufflebeam, Ina Mullis,
Eugene Johnson, Robert Linn

P.P.S. Attached are two Tables to replace earlier Tables 2A and 25. See the footnotes to explain the specific changes that were made.

Table 2A

Summary of Grade 8 Achievement Levels
(Excluding EST and HOTS Items)

		- Achievement Level -								
Item Ratings	N	Basic			Proficient			Advanced		
		Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
1st	22	70.1	14.0	68.8	87.1	9.6	88.0	95.2	4.5	96.3
2nd	22	71.5	16.3	72.7	86.0	9.4	87.5	93.8	3.9	91.2
3rd	22	70.6	14.1	71.1	86.7	9.9	86.3	94.9	4.5	95.8
4th	19	68.9	13.0	69.2	85.1	9.5	86.2	93.9	5.5	94.9
Final[1]	18	64.1	10.5	60.0	81.3	6.4	80.0[2]	91.8	3.2	92.0

[1] Overall ratings based upon the total pool of items.

[2] Final Round Proficient Median was corrected on 12/17/90. The correct number is 80.0, not 81.0.

Table 25

Summary of Participants' Five Sets of Achievement Levels
(Grade 8, 22 Participants, Excluding EST and HOTS Items)

		Basic					Proficient					Advanced				
ID	ED[1]	1	2	3	4	5	1	2	3	4	5[2]	1	2	3	4	5
0808	1	86	89	85	--	--	91	91	92	--	--	96	91	96	--	--
0802	1	58	60	60	53	60	75	77	77	75	76	90	92	92	87	90
0811	1	76	84	84	81	60	99	99	99	96	75	99	100	100	98	90
0815	1	91	87	86	75	66	99	89	97	94	83	100	90	99	99	93
0827	2	69	79	77	--	--	97	97	97	--	--	100	100	100	--	--
0821	1	64	72	81	84	85	88	97	99	97	92	99	100	100	100	97
0806	1	81	79	83	66	60	98	98	98	91	80	100	99	99	97	90
0820	1	76	90	76	61	60	89	90	89	82	80	96	90	97	95	93
0812	1	91	93	92	92	88	99	100	100	98	95	100	100	100	100	99
0816	1	71	74	76	60	60	81	84	86	72	79	93	95	97	86	91
0825	1	76	89	86	78	65	84	94	93	87	83	92	95	97	94	90
0803	1	53	54	54	73	60	79	73	73	83	78	86	91	90	95	92
0828	1	42	45	46	48	50	88	87	87	80	80	96	95	95	93	90
0810	2	65	87	59	45	50	78	88	73	63	70	89	90	84	79	85
0822	1	94	98	91	85	80	99	91	99	95	92	100	91	100	98	94
0823	2	67	67	66	74	64	82	82	82	86	84	92	92	91	94	94
0807	1	58	58	57	56	55	78	78	76	73	75	90	90	89	88	88
0801	1	86	70	65	78	65	93	83	78	89	82	100	97	93	96	92
0826	1	68	44	58	64	60	93	72	83	87	80	99	92	94	97	92
0824	1	53	56	58	67	--	72	73	75	84	--	90	91	92	93	--
0805	1	64	52	54	--	--	87	77	77	--	--	96	92	92	--	--
0809	1	54	54	59	69	65	75	75	77	86	80	90	90	90	96	92
Mean		70.1	71.5	70.6	68.9	64.1	87.1	86.0	86.7	85.1	81.3	95.2	93.8	94.9	93.9	91.8
SD		14.0	16.3	14.1	13.0	10.5	9.6	9.4	9.9	9.5	6.4	4.5	3.9	4.5	5.5	3.2
Median		68.8	72.7	71.1	69.2	60.0	88.0	87.5	86.3	86.2	80.0[2]	96.3	91.2	95.8	94.9	92.0

[1] Educator: 1=Yes; 2=No

[2] Corrected the median for Proficient Students on Round 5 on 12/17/90: 81.0 becomes 80.0.

Appendix H

Panelists for Replication/Validation

Appendix H

Panelists for Replication/Validation

Field Test

Semere Ambaya	Dunbar High School, Washington, D.C.
Ethylene Baker	J. F. Cook Elementary School, Washington, D.C.
Rebecca Barnes	James Madison High School, Vienna, VA
Madelyn Blanding	Gwynn Park High School, Clinton, MD
Jane Bolter	Lanier Intermediate School, Fairfax, VA
Joan Burks	Watkins Mill High School, Gaithersburg, MD
Nancy Carlson	Crossfield Elementary School, Herndon, VA
Jeffrey Choppin	Jefferson Junior High School, Washington, D.C.
Shirley Christman	South Lakes High School, Reston, VA
Bertha Clarke	Frances Scott Key, Middle School, District Heights, MD
Elaine Clarke	Pyle Middle School, Bethesda, MD
Pearl Flowers	Quince Orchard High School, Gaithersburg, MD
Beryl Jackson	Instructional Service Center, Washington, D.C.
Fay Jackson	Greenbelt Middle School, Greenbelt, MD
Zenobia Justice	Murch Elementary School, Washington, D.C.
Linda Kostenbader	Terra Centre Elementary School, Burke, VA
Gerry May	Redland Middle School, Rockville, MD
Sally Roth	Key Intermediate School, Springfield, VA
Fred Sanford	High Point High School, Mitchellville, MD
Debbie Stone	Laurel Elementary School, Laurel, MD

Barbara Williams	Montgomery Knolls Elementary, Silver Spring, MD
Jacqueline Williams	Eastern High School, Washington, D.C.
Lynn Wittington	Skyline Elementary School, Upper Marlboro, MD

California

Harold Asturias	Los Angeles Unified School District, Los Angeles, CA
Cheryl Avalos	Gladstone High School, Covina, CA
Steve Balok	Sinaloa High School, Novato, CA
Pam Beck	Fresno Unified School District, Fresno, CA
Jerry Bernhardt	Amy Blanc Elementary, Fairfield, CA
Lloyd Berriman	Los Angeles Unified School District, Long Beach, CA
Kathy Blackwood	Los Angeles Unified School District, Long Beach, CA
Beverly Braxton	Willard Jr. High School, Berkeley, CA
Carol Brooks	Oakland Unified School District, Oakland, CA
Jeanette Burds	Sutter Jr. High School, Canoga Park, CA
Carol Buss	Irvine High School, Irving, CA
Dianne Camacho	Warren High School, Downey, CA
Marie Carrick	Sharp Park School, Pacifica, CA
Amarjit Chadda	Los Altos High School, Los Altos, CA
William Collins	Jameslick High School, San Jose, CA
Cathy Crowell	San Jose Unified School District, San Jose, CA
Margaret DeArmond	East High School, Bakersfield, CA
Marilyn Dickens	Ukiah Unified School District, Ukiah, CA
Linda Dritsas	Jameslick High School, San Jose, CA
Joe Duardo	Member, Board of Education, Whittier, CA
Jim Feenstra	Mt. Diablo School District, Mt. Diablo, CA
Lee Gotcher	Warner Middle School, Westminster, CA
Owen Griffith	Member, Board of Education, Torrance, CA

Rosalyn Haberkern	Crocker Highlands Elementary, Oakland, CA
Audrey Hanson	Member, Board of Education, Burbank, CA
Linda Haysom	Garden Grove Unified School District, Garden Grove, CA
Hal Hendrickson	Member, Board of Education, Morgan Hills, CA
Valerie Henry	Sierra Vista Middle School, Irvine, CA
Christine Hiroshima	Department of Integration, J. Smith Center, San Francisco, CA
David L. Hughes	River Delta Unified School District, Clarksburg, CA
Joyce Ireland	Santa Ana Unified School District, Santa Ana, CA
Joy Kelly-McBurney	Temecula Valley Unified School District, Temecula Valley, CA
Dorothy Kirk	Sumerset Sr. High School
Joyce Kirsch	Los Angeles Unified School District, North Hollywood, CA
Drew Kravin	Albany Unified School District, Albany, CA
Ted Lobman	Stuart Foundation, San Francisco, CA
Jamila Makini	Emery High School, Emeryville, CA
Feliciano Mendoza	Los Angeles Unified School District, Huntington Park, CA
Teferi Messeret	Fernbacon Middle School, Sacramento, CA
Clarita Montalbon	Jurupa Valley High School, Riverside, CA
Sara Munshin	Roosevelt High School, Los Angeles, CA
Juanita Ortman	Pasadena, CA
Jackie Palmer	Middleton Street School, Huntington Park, CA
Louisa Perez	Member, Board of Trustees, Sacramento, CA
Garlyn Peterson	Novato Unified School District, Novato, CA
Serena Pon	Oakland Unified School District, Oakland, CA
Jenny Reid	Riverside Unified School District, Riverside, CA

Joan Robinson	Newport Mesa Unified School District, Costa Mesa, CA
Karen Rogge	California P.T.A., Oakland, CA
Joel Roszell	Long Beach Unified School District, Long Beach, CA
Nancy Schager	Ocean View Unified School District, Huntington Beach, CA
Richard Shiers	Lompoc Unified School District, Lompoc, CA
Sharon Stuart	Simi Valley Unified School District, Simi Valley, CA
Karl Ting	Morgan Hill Unified School District, Morgan Hill, CA
Lisa Usher	Audum Jr. High School, Los Angeles, CA

Connecticut

Patricia Banning	Kramer Middle School, Willimantic, CT
Oliver Barton	High School in the Community, New Haven, CT
Jerry Bencivenga	State Department of Education, Middletown, CT
Katherine Bishop	Daisy Ingraham School, Westbrook, CT
Sandra Brandt	Pomfret Community School, Pomfret Center, CT
Jeanne Cavallaro	Milford, CT
Sandra Coelho	E. Windsor Intermediate School, Bethany, CT
Sharon Cooley	Lincoln School, New Britain, CT
Thomas Day	Wallingford, CT
Gail Dichiara	Westbrook High School, Westbrook, CT
Robert Dion	Staples High School, Westport, CT
Tony Ditrio	Norwalk Public Schools, Norwalk, CT
Winifred Dixon	Dwight School, New Haven, CT
Diane Dzikiewicz	O'Brien School, East Hartford, CT
Debra Feldman	Hamilton Avenue School, Greenwich, CT
Roger Fiondella	Fairfield Public School, Fairfield, CT
Frederick Fitzgerald	E. Hartford Middle School, East Hartford, CT
Jane Furey	Searles Middle School, Great Barrington, MA
Dennis Gannon	Francis T. Maloney H.S., Meriden, CT
Heather Giancola	Springdale Elementary School, Stamford, CT
Dennis Grant	Windsor High School, Windsor, CT
Margaret Guaneri	Griswold Elementary School, Jewett City, CT
Debra Isenstein	Dunbar School, Bridgeport, CT

Karen Jones	Springfield, MA
Marshall Kelly	New Haven, CT
Katherine Kocher	Naramake Elementary School, Norwalk, CT
Henry Kopij	Montville High School, Oakdale, CT
Bernadine Krawczyk	Wooster Middle School, Stratford, CT
James Landherr	E. Hartford High School, Preston City, CT
Dan Lawler	West Hartford, CT
Jeffrey Leo	Pomfret Community School, Pomfret Center, CT
Edward Lestinski	Vogel Elementary School, Torrington, CT
Patricia Llodra	Northwestern Regional H.S., Winstead, CT
Sue Marchitto	Regional Water Authority, New Haven, CT
Patsy Mayo	Hill Central School, New Haven, CT
Peg McDonald	ASA, The Pension Service, New Haven, CT
Rufus Morton	Bristol Eastern H.S., Bristol, CT
Joanna Panning	Middletown H.S., Middletown, CT
Maryann Papa	Conrad High School, West Hartford, CT
Jorge Pezo	Harding High School, Bridgeport, CT
Helen Prescott	Ashford Elementary School, Ashford, CT
Debbie Richardson	Carmen Arace School, Bloomfield, CT
Norman Ricker	New Canaan High School, New Canaan, CT
Kenny Sherrick	Berlin High School, Berlin, CT
Mari Smith	Hartford, CT
Beverly Stern	Hillhouse High School, New Haven, CT
James Thomas	Lennox, MA

Frank Tomaino	Newtown High School, Sandy Hook, CT
Lawrence Tripp	New Milford, CT
Lester Turner	James Hillhouse H.S., New Haven, CT
Janice Vuolo	Cheshire, CT
Darlene Wallin	New Milford, CT
Peter Warren	Amity Regional Jr. High School, Bethany, CT
Ellie Zaloski	New Milford, CT

Florida

Susan Atteridge	Director, Corporate Affairs, AT&T, Miami, FL
James B. Bailey	Zephyrhills High School, Zephyrhills, FL
Marsha Berdit	Alfred duPont Jr. High, Jacksonville, FL
Ann Blomquist	Boone High School, Orlando, FL
Richard Bradley	Van Buren Jr. High, Tampa, FL
Mary Brinson	Winter Park High School, Winter Park, FL
Patricia Carroll	Vero Beach Jr. High, Vero Beach, FL
Lou Cerreta	Lewis Elementary School, Temple Terrace, FL
Shirley G. Cherry	R.B. Steward Middle School, Zephyrhills, FL
Wendy D'Agostino	Union Park Middle School, Orlando, FL
Elaine Dutton	St. Andrew's School, Ft. Pierce, FL
Gwinetta Evans	Bay Haven Elementary, Sarasota, FL
Elisie Flores	Melrose Elementary, Miami, FL
Georgia Forbes	Edison Middle School, Miami, FL
Steve Frielander	Richards High School, Tallahassee, FL
Nelson Garcia	Jose Mart Middle School, Hialeah, FL
Shirley Hall	Miami Center Sr. High, Miami, FL
Rosa B. Hill	Pasco Middle School, Dade County, FL
Steve Horton	Megis Middle School, Shalimar, FL
Alice Hough	Wright Elementary, Miami, FL
Harrison Howard	Hammocks Middle School, Kendall, FL
Pam Inmann	New Directions High School, Sarasota, FL
Mike Jacobs	Miami Museum of Science, Miami, FL

Gordon James	Pine Villa Elementary, Goulds, FL
Jim Kelly	Venice Area Middle School, Venice, FL
Ramesh Krishnaiyer	Florida Atlantic University, Ft. Lauderdale, FL
Emily Landreth	Godby High School, Tallahassee, FL
Edwina Laymon	Florida PTA, Ft. Myers, FL
Rhesa Marshall	Godby High School, Tallahassee, FL
Mary Ellen Martin	Treasure Island Elementary, Miami Beach, FL
Toy Martinez	Carrollwood Elementary, Tampa, FL
Randall McComas	IBM Corporation, Tempa, FL
Jacqueline Paulk	Bay Haven Elementary, Sarasota, FL
John Pecott	Paxon Middle School, Jacksonville, FL
Beverly Peters	Robinson Sr. High School, Tampa, FL
Evelyn Price	Plant Sr. High School, Tampa, FL
Mary Pritchett	Chamber of Commerce, Tallahassee, FL
Ann Putnam	Ashton Elementary, Sarasota, FL
Ryan Roberts	Seminole Electric, Cooperative, Inc., Tampa, FL
John Sanders	Turkey Creek Jr. High, Plant City, FL
Janet Schacht	Rosewood Elementary, Vero Beach, FL
Michael Shallow	Vero Beach High School, Vero Beach, FL
Ellen Shepherd	Niceville High School, Niceville, FL
Harvey Smerilson	Meadowbrook Middle School, Orlando, FL
Ivy Tubbs	Venice High School, Venice, FL
Lynn Volpe	Bloomingdale Senior High School, Valrico, FL

Richard Westover	Riverview High School, Sarasota, FL
Merlyn Williams	Knights Elementary School, Plant City, FL

Michigan

Gayle Barton	Monroe Elementary School, Wyandotte, MI
Donna Beach	Comstock Middle School, Comstock, MI
Murney Bell	New Baltimore, MI
Ann Beyer	Ann Arbor, MI
Ann Browning	The Upjohn Company, Kalamazoo, MI
Pat Carlson	Fruitport High School, Fruitport, MI
Alice Cole	Highland Park School District, Highland Park, MI
Robert Cook	Tyler Elementary School, Belleville, MI
Cherie Cornick	Roosevelt High School, Wyandotte, MI
Ken DaRos	Woodworth Junior High, Dearborn, MI
Paul Eckhert	Kalamazoo Public Schools, Kalamazoo, MI
Richard Elsholz	Waterford Public Schools, West Bloomfield, MI
Kim Fairchild	North Middle School, Belleville, MI
Janet Fuller	Coldwater Community Schools, Coldwater, MI
Mary Gilkey	Baylor Elementary School, Inkster, MI
Katie Gorignon	River Rouge Public Schools PTA, River Rouge, MI
Spencer Grant, Jr.	Blanchette Junior High, Inkster, MI
Ron Green	Portland Public Schools, Portland, MI
Marilyn Hansbarger	Wacousta School, Eagle, MI
Arthur Harris	Sabbeth Elementary School, River Rogue, MI
Bill Harris	Huron High School, Ann Arbor, MI
Jeanne Herrmann	South Redford School District, Redford, MI

Judy Higbee	Silver Springs Elementary, Northville, MI
Sue Ann Hise-Denk	Connections, Rochester, MI
Laurie Hochrein	Clague Middle School, Ann Arbor, MI
Jan Edward Hulett	Grand Blanc, MI
Deborah Jenkins	Pershing High School, Detroit, MI
Anita Johnston	Napoleon High School, Napoleon, MI
Jean Kelsey	Angell Elementary School, Ann Arbor, MI
Laurie Kohout	Flint Community School, Flint, MI
Linda Kolnowski	East Detroit Public Schools, East Detroit, MI
Debbie Larner	Holt Middle School, Holt, MI
Chris Laske	Meijer, Inc., Grand Rapids, MI
Karen Lauterbach	Gardner S. Wilmington H.S., Gardner, IL
Tom McIntyre	Willow Run Community Schools, Ypsilante, MI
Ken Mass	Minooka High School, Minooka, MI
Warren Matthews	Slauson Middle School, Ann Arbor, MI
Patricia McMann	Roosevelt High School, Wyandotte, MI
Marie Miller	River Rogue High School, River Rogue, MI
James Moser	General Motors, Belleville, MI
Roberta Papora	Ford Elementary School, Ypsilanti, MI
Bill Parish	T.N. Lamb Jr. High School, Burton, MI
David Powell	East Detroit Public Schools, East Detroit, MI
James Rossi	Traverse City High School, Traverse City, MI
Gene Rummell	Michigan National Bank, Lansing, MI
Robin Rutz	Bach Open Elementary School, Ann Arbor, MI

Steve Saliba	Braidwood Elementary School, Braidwood, MI
Jane Schleeter	Plainfield Jr. High School, Plainfield, IL
Frances Scott	Kaiser Elementary School, Ypsilanti, MI
Lynn Serenson	Novi Middle School, Novi, MI
Nancy Skwarczynski	Minooka Jr. High School, Minooka, IL
Karma Storm	Pinecrest Elementary School, East Lansing, MI
Beverly Tyler	Ardis Elementary School, Ypsilanti, MI
Nancy Varner	Detroit Public Schools, Detroit, MI
Cheryl Vaughn	Ferndale School District, Ferndale, MI
Cathy Walter	Rawsonville Elementary School, Ypsilanti, MI
Sue Wright	Consumers Power, Essexville, MI
William York	Holt High School, Holt, MI

Appendix I

Summary of Validation/Replication Achievement Level Setting Data

Table 40. **Expected Proportion-Correct** Scores for the Basic, Proficient, and **Advanced Levels** (Grade = 4, Block[1] = 3, Judges = 30)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	68.5	73.7	86.0	88.8	97.3	97.7
2	59.1	45.1	80.6	70.9	94.8	87.4
3	71.6	71.0	86.6	87.9	96.8	96.6
4	76.7	80.6	90.4	93.6	97.9	98.4
5	61.8	59.6	80.9	81.5	93.0	95.0
6	48.6	41.8	70.8	65.3	87.0	84.1
7	58.8	54.2	79.1	75.8	93.8	90.3
8	65.8	68.3	84.9	85.9	96.1	96.2
9	56.5	59.1	76.3	79.3	92.2	92.3
10	50.5	45.4	72.4	68.8	88.0	86.5
11	63.8	65.9	84.2	84.3	95.4	95.7
12	56.9	55.9	78.8	77.0	92.6	92.2
13	53.5	48.0	76.0	73.1	91.4	88.5
14	57.8	61.7	80.6	81.7	94.7	95.1
15	50.9	41.3	75.0	67.0	90.6	84.7
16	56.7	55.6	79.6	78.6	93.0	91.7
17	45.4	34.7	71.8	63.7	89.4	82.3
18	42.1	43.3	68.2	70.7	87.8	87.4
19	41.5	43.0	69.5	69.9	86.4	87.6
$\bar{x}$	57.2	55.2	78.5	77.0	92.5	91.0
SD	9.6	12.8	6.3	8.6	3.6	5.0

The tables that follow summarize the item level ratings by the judges on a block-by-block basis. There are 7 blocks of items for each grade level. The number of judges per block varies depending upon the total number of judges present at the grade-level sessions and the specific student booklets distributed at the sessions.

---

[1] The total number of items in a student booklet are divided into blocks, consisting of about 20 items. Each Student booklet contains 3 blocks. In 1990, the cognitive item block were numbered from 3 to 9. Background questions were numbered 1 and 2; and special study items (HOTS and EST) were numbered 10 to 12.

Table 41. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 4, Judges = 25)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	64.7	68.6	85.0	88.1	96.6	97.3
2	60.6	62.4	82.8	82.5	96.0	94.4
3	40.7	47.5	66.8	70.7	84.0	88.2
4	43.0	36.6	71.8	66.7	90.2	84.5
5	29.4	27.4	59.2	55.4	82.6	80.2
6	34.2	27.7	65.6	57.7	85.3	78.3
7	26.5	21.2	51.4	44.0	73.8	67.9
8	30.4	25.1	61.9	54.0	83.5	77.2
9	43.3	30.2	69.8	56.9	89.0	79.2
10	23.6	19.6	57.4	51.7	82.2	74.3
11	34.0	27.7	65.4	58.6	86.2	80.4
12	17.0	13.8	43.1	37.8	65.6	60.3
13	20.6	14.9	50.8	41.7	76.2	68.3
14	25.6	16.2	58.0	43.3	82.5	68.1
$\bar{x}$	35.3	31.4	63.5	57.8	83.8	78.5
SD	14.1	17.0	11.7	14.9	8.2	10.4

Table 42. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 5, Judges = 30)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	58.3	49.8	78.3	70.5	95.3	90.1
2	34.5	18.5	56.5	41.4	80.3	63.7
3	39.1	29.7	64.3	52.9	84.3	76.2
4	54.3	48.6	78.2	73.2	94.7	91.1
5	49.2	35.7	73.9	61.9	92.0	84.5
6	41.9	41.1	70.2	68.2	89.5	88.7
7	47.6	37.5	73.7	62.0	92.7	85.5
8	36.7	30.7	67.1	58.7	86.4	80.3
9	42.5	33.3	69.3	58.4	89.0	81.8
10	62.2	63.5	83.7	84.2	96.5	95.7
11	25.2	20.6	55.1	46.7	76.6	69.0
$\bar{x}$	44.7	37.2	70.0	61.6	88.8	82.4
SD	11.0	13.2	8.9	12.2	6.4	9.7

Table 43. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 6, Judges = 26)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	44.5	43.5	70.3	68.8	90.8	90.4
2	39.5	28.2	66.0	55.8	85.8	78.3
3	56.7	60.4	80.4	81.5	95.6	93.6
4	50.0	28.1	75.6	55.0	93.0	77.5
5	45.7	40.8	73.6	68.8	89.7	86.6
6	41.4	36.5	73.6	66.7	90.6	85.9
7	48.3	30.0	76.4	59.0	91.7	81.7
8	35.2	33.7	63.9	62.3	86.3	84.6
9	45.3	35.0	70.8	64.3	88.6	85.1
10	52.4	43.8	78.1	68.7	93.8	88.7
11	43.4	42.2	71.3	70.2	89.9	87.6
12	32.0	27.7	65.0	59.4	86.7	81.8
13	40.3	28.2	70.3	57.3	87.5	78.8
14	27.3	18.4	57.9	45.4	78.3	67.7
15	36.0	23.3	65.4	52.2	84.7	75.0
16	37.7	22.1	66.4	52.7	83.7	73.3
17	40.4	23.7	69.4	53.8	90.1	76.5
$\bar{x}$	42.1	33.3	70.3	61.3	88.6	81.9
SD	7.6	10.5	5.8	8.9	4.2	6.8

NOTE: Grade = 4, Block = 6, Judges = 26

Table 44. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 7, Judges = 22)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	48.7	63.1	69.3	80.0	88.8	95.8
2	45.2	57.6	69.0	77.6	88.4	94.3
3	55.4	60.9	75.4	80.3	91.0	95.6
4	49.8	54.1	71.5	74.3	89.8	92.1
5	53.5	53.4	75.2	75.4	91.3	92.1
6	39.8	38.9	63.9	65.5	84.5	84.5
7	47.9	48.2	72.6	73.8	92.3	91.9
8	40.1	34.5	67.0	62.8	88.5	84.1
9	39.8	42.4	65.0	68.7	86.6	88.1
10	41.9	40.2	67.5	67.0	88.4	87.3
11	34.2	36.5	60.0	63.4	79.5	82.8
12	50.8	40.5	73.1	67.0	92.3	86.5
13	39.0	37.2	64.9	63.0	86.4	84.8
14	48.7	45.3	72.3	71.5	90.4	89.5
15	25.4	22.9	50.4	47.2	73.6	69.5
16	37.2	31.9	60.9	57.0	82.0	79.7
17	35.4	35.5	62.5	61.9	83.9	83.5
18	34.8	35.4	61.0	61.3	82.5	81.7
$\bar{X}$	42.6	43.3	66.8	67.7	86.7	86.9
SD	7.9	10.9	6.4	8.6	4.9	6.5

Table 45. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 8, Judges = 33)

Item	Basic		Proficient		Advanced	
	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	66.7	73.0	83.5	87.5	97.1	97.7
2	58.5	59.5	79.4	80.4	93.2	93.3
3	53.5	57.2	75.1	77.3	90.8	92.4
4	44.5	41.1	68.8	66.1	90.1	85.3
5	45.4	40.4	69.2	65.2	88.5	84.8
6	42.8	31.7	67.9	59.2	87.7	79.0
7	53.8	53.8	74.5	75.3	91.2	90.2
8	48.5	39.4	70.6	62.8	89.3	84.9
9	53.2	47.5	76.0	69.1	91.9	88.5
10	45.6	39.5	69.6	59.8	89.0	81.1
11	48.5	44.9	72.1	66.9	89.5	87.7
12	60.4	54.7	79.7	75.3	94.2	90.8
13	50.7	46.4	73.2	69.2	90.8	86.1
14	46.5	40.9	70.3	62.4	88.5	82.9
15	37.3	31.8	62.4	56.0	81.4	77.8
$\bar{x}$	50.4	46.8	72.8	68.8	90.2	86.8
SD	7.5	11.2	5.4	8.8	3.5	5.5

Table 46. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 4, Block = 9, Judges = 29)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	63.1	66.7	84.0	86.6	96.3	97.7
2	61.5	58.9	82.4	81.3	97.1	95.3
3	63.3	58.9	83.1	78.7	94.5	92.6
4	34.4	38.4	63.6	66.5	84.9	86.4
5	46.8	41.9	73.9	68.0	90.5	85.5
6	46.8	42.7	73.0	68.6	90.4	87.5
7	48.8	44.1	73.3	70.0	91.8	89.1
8	60.0	47.7	82.9	72.6	95.9	90.2
9	49.6	45.1	73.5	69.9	92.3	89.4
10	41.4	33.6	71.2	62.3	90.4	83.3
11	25.2	17.0	53.7	40.2	79.8	66.3
12	35.3	19.4	60.9	45.0	85.9	66.8
13	50.7	34.4	72.3	58.3	89.7	77.1
14	51.1	35.4	75.1	59.1	91.7	78.7
15	48.0	30.7	73.1	58.3	88.6	78.3
$\overline{x}$	48.4	41.0	73.1	65.7	90.7	84.3
SD						

Table 47. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 3, Judges = 27)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	74.7	72.1	87.9	86.0	96.3	95.6
2	67.7	70.6	85.6	85.8	95.5	95.9
3	63.7	67.6	86.3	85.5	96.1	96.1
4	68.4	73.8	87.3	88.6	97.1	97.7
5	63.3	62.5	84.9	82.7	95.3	94.1
6	55.0	56.5	78.1	80.0	92.7	93.0
7	66.1	61.7	85.4	83.8	96.9	95.9
8	58.1	66.2	78.1	84.2	93.1	94.6
9	53.1	65.8	77.7	83.7	93.1	95.6
10	55.5	59.5	79.7	82.2	93.8	94.2
11	48.6	44.0	75.4	69.4	92.5	89.8
12	35.5	42.3	62.4	67.3	84.7	86.3
13	55.2	47.4	80.0	73.7	94.6	91.8
14	50.3	57.1	78.5	82.0	91.7	94.1
15	41.8	44.9	71.7	73.6	90.3	90.6
16	41.1	39.5	68.9	66.7	88.4	86.4
17	52.3	50.1	79.0	73.9	95.3	92.8
18	43.4	42.6	73.5	73.5	91.1	90.0
19	53.4	49.9	79.7	75.8	95.7	93.7
20	40.0	34.7	70.7	62.3	90.0	85.2
21	38.3	36.8	68.4	65.4	89.1	87.8
22	52.0	43.2	75.5	68.6	92.8	88.0
23	32.1	27.1	64.1	57.6	86.7	82.3
$\overline{x}$	52.6	52.9	77.3	76.2	92.7	91.8
SD	11.4	13.2	7.2	8.8	3.3	4.4

Table 48. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 4, Judges = 31)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	79.3	81.3	91.0	92.1	97.1	98.2
2	76.6	77.5	88.2	89.1	96.8	97.2
3	49.1	56.2	69.5	76.5	87.5	90.4
4	60.8	61.3	80.7	81.4	94.7	95.1
5	43.9	47.4	71.6	72.3	91.0	91.2
6	53.2	56.1	76.8	78.7	93.2	93.9
7	39.5	37.7	67.1	65.2	86.5	85.3
8	50.4	50.0	75.3	73.8	92.2	91.0
9	58.3	57.8	78.1	77.4	95.2	94.6
10	45.0	44.8	72.8	71.9	89.0	89.0
11	55.1	55.7	80.2	80.0	93.7	94.2
12	35.9	33.8	66.1	62.9	86.8	85.4
13	39.3	32.0	67.9	62.3	88.1	84.9
14	42.9	40.8	72.9	70.7	90.5	89.8
15	31.3	22.3	62.0	51.9	84.8	79.4
16	30.1	24.8	60.4	52.5	83.0	77.4
17	29.2	25.3	60.6	55.8	83.7	80.9
18	27.5	23.4	56.1	50.7	78.5	74.7
19	24.9	20.9	54.8	48.3	79.1	74.9
20	23.4	20.6	52.5	49.9	82.2	79.3
21	18.8	16.6	49.6	45.9	77.7	74.9
$\overline{x}$	43.5	42.2	69.2	67.1	88.2	86.7
SD	16.5	18.9	11.3	14.0	6.0	7.8

Table 49. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 5, Judges = 31)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	72.1	67.6	88.3	87.1	98.7	97.6
2	63.5	57.9	84.3	80.6	97.3	95.6
3	52.3	43.5	75.0	69.8	90.9	88.3
4	73.2	63.9	89.6	85.0	97.8	95.6
5	72.1	63.7	88.8	85.8	98.3	96.1
6	65.3	67.6	85.3	86.0	95.8	96.4
7	60.1	52.8	81.9	77.3	96.0	93.5
8	49.3	42.7	74.8	69.0	90.6	88.0
9	56.9	51.6	80.0	78.0	95.1	93.4
10	35.1	24.2	62.3	51.6	84.3	76.2
11	39.7	31.8	65.9	59.5	88.4	83.3
12	34.9	31.4	67.3	62.5	88.7	86.1
13	50.5	42.7	75.5	68.4	90.6	86.9
14	42.2	32.9	71.4	63.7	90.4	85.3
15	42.6	36.6	71.1	65.2	87.9	86.0
16	43.6	38.1	71.0	66.1	87.2	85.7
$\bar{x}$	53.3	46.8	77.0	72.2	92.4	89.6
SD	13.2	14.2	8.7	10.8	4.6	6.0

Table 50. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 6, Judges = 28)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	65.7	69.1	84.8	87.0	95.7	96.5
2	62.2	61.1	82.5	82.0	94.2	94.3
3	5.1	73.5	83.8	88.3	94.9	97.1
4	56.0	41.1	77.2	62.7	91.0	81.3
5	61.6	70.0	81.7	87.4	95.9	97.5
6	57.4	58.3	80.1	81.0	94.6	95.1
7	59.2	53.3	81.3	76.0	94.4	92.6
8	46.8	43.6	72.4	67.8	91.5	89.7
9	55.3	55.3	78.8	78.1	93.1	94.0
10	62.0	59.2	83.3	80.0	95.6	95.6
11	54.8	57.9	79.8	79.3	94.2	94.5
12	55.3	53.3	78.1	76.6	93.7	92.5
13	54.0	52.1	76.9	74.6	93.0	92.2
14	41.8	39.8	67.9	62.9	86.8	85.2
15	45.9	44.9	72.5	70.2	90.3	90.4
16	52.0	53.1	75.8	77.1	94.0	94.8
17	56.3	50.0	79.5	73.6	94.8	92.7
18	35.3	28.8	63.8	53.3	84.0	79.5
19	46.8	44.4	72.5	69.3	91.1	89.6
20	44.3	40.0	66.9	60.0	86.8	83.9
21	42.0	38.8	69.9	64.8	90.7	87.8
$\overline{X}$	53.3	51.8	76.6	73.9	92.4	91.3
SD	8.4	11.4	6.0	9.5	3.2	5.1

Table 51. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 7, Judges = 28)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	56.4	51.3	81.6	77.3	96.3	94.9
2	45.6	38.6	77.8	71.3	94.3	91.5
3	47.8	46.3	78.9	74.2	94.1	91.1
4	64.6	73.2	86.2	88.8	98.4	98.3
5	52.7	44.5	77.8	70.5	94.6	91.8
6	35.2	34.3	63.3	60.7	87.8	86.4
7	44.2	39.2	73.2	66.8	92.5	90.1
8	40.4	41.1	71.3	70.3	91.3	91.5
9	41.3	38.4	71.3	67.0	92.3	89.3
10	55.5	52.8	79.1	74.1	94.6	89.1
11	39.3	31.1	67.0	58.9	91.0	84.8
12	30.3	21.6	61.7	49.6	83.9	77.0
13	28.1	22.3	58.3	49.3	83.7	77.5
14	52.3	40.2	79.3	70.5	95.1	89.2
15	40.8	29.1	71.1	60.1	92.9	87.1
16	30.0	25.7	59.9	55.1	83.1	80.2
17	30.5	20.0	58.8	48.4	80.8	73.6
18	40.0	35.7	68.4	63.8	90.1	85.7
$\overline{x}$	43.1	38.1	71.4	65.4	90.9	87.1
SD	10.3	13.1	8.5	10.7	5.1	6.5

Table 52. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 8, Judges = 25)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	59.6	74.3	83.0	89.5	95.1	97.8
2	57.8	68.6	80.5	85.9	94.8	96.4
3	55.8	52.5	79.1	77.3	93.6	91.7
4	45.9	41.3	71.3	66.3	89.4	86.8
5	47.8	43.4	75.7	72.2	92.6	89.6
6	42.2	37.2	70.0	64.7	87.8	83.4
7	43.9	38.8	72.8	67.3	90.0	88.2
8	32.6	29.9	62.8	58.4	84.2	80.9
9	36.0	29.6	65.8	56.4	86.4	77.2
10	33.5	24.7	64.6	54.0	84.3	76.2
11	24.1	18.2	54.7	46.2	77.8	70.5
12	74.3	76.4	89.1	90.7	98.5	99.1
13	68.9	66.7	86.1	84.7	96.9	96.8
14	64.9	64.5	85.5	84.3	95.9	96.2
15	54.8	49.6	78.6	73.8	93.7	89.9
16	30.9	25.2	59.4	50.8	78.2	70.8
17	44.5	38.5	72.5	63.2	90.6	84.5
18	34.4	26.8	63.9	56.4	84.1	76.4
$\overline{x}$	47.3	44.8	73.1	69.0	89.7	86.2
SD	14.2	18.5	9.9	14.0	6.1	9.3

Table 53. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 8, Block = 9, Judges = 28)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	66.4	75.7	84.8	90.8	95.6	98.5
2	60.5	69.5	82.0	85.8	95.4	97.0
3	58.7	62.1	80.5	82.1	93.9	94.6
4	40.8	34.6	71.8	65.9	89.3	86.5
5	42.3	45.5	70.6	71.8	87.7	88.6
6	53.4	55.3	78.5	79.1	93.8	94.0
7	36.3	39.1	65.2	67.0	86.6	89.7
8	53.8	48.2	76.4	73.4	92.6	92.4
9	49.6	43.8	77.1	72.4	94.7	91.3
10	47.5	42.0	77.1	71.3	93.9	91.8
11	46.8	43.6	75.3	72.0	93.3	91.5
12	32.8	26.8	65.2	59.8	87.2	85.1
13	51.1	44.3	77.6	73.2	94.5	92.8
14	35.7	31.6	65.5	60.2	87.8	85.5
15	56.8	47.7	81.1	72.6	95.0	93.9
16	47.7	47.1	74.8	73.6	91.8	91.6
17	40.9	27.8	69.3	58.0	90.0	84.0
18	31.1	23.6	64.6	56.4	85.7	81.4
19	25.9	18.8	59.6	52.0	83.8	80.0
20	34.2	28.0	68.3	62.3	89.3	85.8
$\bar{x}$	45.6	42.8	73.3	70.0	91.1	89.8
SD	10.9	15.0	6.9	10.0	3.7	5.0

Table 54. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 3, Judges = 32)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	67.2	75.2	86.5	89.9	97.7	98.7
2	63.0	68.3	85.4	87.4	96.6	97.1
3	68.1	65.0	87.6	85.0	97.8	97.3
4	70.8	74.3	89.0	89.7	98.2	98.4
5	49.8	63.1	81.7	87.0	96.9	97.8
6	57.0	65.3	88.2	89.6	98.1	98.6
7	67.3	63.4	89.0	86.6	98.1	97.7
8	65.5	65.5	86.8	85.7	97.3	97.1
9	39.7	51.3	75.9	81.9	95.9	97.5
10	55.2	57.3	83.6	83.5	96.4	96.2
11	58.1	55.6	84.2	80.4	95.4	94.5
12	54.7	55.6	82.4	80.9	96.1	95.8
13	39.1	37.8	78.5	75.3	95.6	93.7
14	53.1	49.4	81.7	79.2	95.8	95.3
15	34.7	38.9	72.4	75.2	92.5	93.3
16	36.3	34.2	73.8	71.2	94.4	92.3
17	34.5	33.1	67.0	64.7	88.0	86.6
18	31.8	30.1	70.8	67.0	94.3	92.0
19	37.8	31.4	72.3	65.0	91.8	88.6
20	27.7	17.6	60.4	50.1	86.0	78.5
21	16.4	14.6	50.5	46.4	85.0	82.7
22	41.1	38.1	74.1	69.5	91.8	90.1
23	34.7	23.9	68.4	58.8	91.3	85.3
$\bar{x}$	48.0	48.2	77.8	76.1	94.4	93.3
SD	15.3	18.3	10.0	12.6	3.8	5.6

Table 55. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 4, Judges = 31)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	87.3	89.3	95.6	95.7	98.9	99.1
2	85.4	86.2	94.4	94.5	98.3	98.4
3	56.7	66.0	76.2	82.4	92.2	95.5
4	72.2	75.1	88.4	88.5	97.6	98.2
5	59.6	62.5	81.3	82.5	95.4	94.7
6	68.9	73.3	87.3	88.7	97.8	97.7
7	45.7	49.4	70.4	73.4	88.0	88.8
8	60.7	60.8	84.4	83.0	96.3	95.6
9	73.1	74.4	89.2	88.7	97.5	96.9
10	60.2	65.2	80.7	83.1	93.9	94.8
11	67.3	71.3	87.5	87.7	97.3	96.9
12	49.9	49.8	76.9	76.0	92.9	92.2
13	51.7	45.7	79.4	74.1	92.7	89.7
14	54.9	57.8	82.3	84.0	94.8	95.3
15	45.6	44.6	78.8	77.3	93.2	93.1
16	49.9	40.2	74.6	66.1	89.1	84.6
17	38.2	40.1	72.9	73.0	92.3	92.3
18	40.3	31.9	71.7	65.5	90.7	86.6
19	35.8	26.9	67.4	59.5	86.7	80.6
20	31.5	31.4	68.6	69.1	90.0	89.1
21	13.4	10.8	48.8	43.3	79.5	75.8
22	10.5	6.6	36.2	28.8	73.6	67.5
$\bar{x}$	52.7	52.7	77.0	75.7	92.2	91.1
SD	19.8	22.5	13.8	16.1	6.2	8.0

Table 56. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 5, Judges = 29)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	61.4	54.7	82.4	78.7	97.3	95.1
2	53.3	51.9	76.3	76.2	93.8	92.8
3	66.0	66.9	84.5	85.8	97.3	95.7
4	27.9	17.3	59.2	44.9	82.0	67.9
5	22.4	15.9	53.4	45.3	81.2	72.8
6	68.0	70.2	84.5	86.2	96.6	96.8
7	71.4	69.7	88.8	87.9	98.6	97.2
8	57.1	50.7	78.1	73.5	91.6	89.0
9	66.8	64.6	86.4	84.9	97.7	96.6
10	37.8	28.6	65.8	55.3	85.9	77.9
11	48.4	35.2	72.0	61.2	90.9	82.8
12	43.0	44.9	75.2	74.0	93.7	93.1
13	57.4	55.7	79.2	78.1	93.4	92.5
14	43.2	41.8	75.3	75.6	93.9	93.1
15	43.3	32.0	75.3	64.6	93.2	85.4
16	29.1	26.4	64.0	61.0	88.1	84.4
17	21.2	15.0	56.3	44.1	81.5	73.6
$\overline{x}$	48.1	43.6	73.9	69.2	91.6	87.5
SD	16.4	19.0	10.7	15.0	5.8	9.4

Table 57. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 6, Judges = 29)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	77.0	81.5	89.9	92.1	98.0	98.7
2	72.8	75.9	89.2	91.0	97.9	98.3
3	68.2	63.3	87.5	85.2	96.7	95.2
4	70.7	73.8	90.2	91.6	97.5	98.2
5	59.8	55.0	84.3	82.0	95.7	93.9
6	33.5	28.9	70.8	65.8	91.2	89.2
7	49.2	44.0	75.8	71.1	91.3	89.3
8	31.1	26.1	70.8	64.4	91.6	88.9
9	50.5	49.5	73.2	71.5	89.0	88.2
10	54.8	50.5	77.9	74.7	91.8	88.6
11	31.6	25.1	68.2	59.5	86.9	82.8
12	50.3	48.3	79.5	78.2	93.7	92.7
13	23.0	19.0	58.2	51.6	80.7	77.0
14	51.2	51.0	84.3	83.8	96.5	96.3
15	32.4	26.5	65.5	58.8	85.4	81.4
16	20.3	15.3	57.4	47.9	80.4	76.2
17	43.6	37.3	72.4	67.2	89.6	85.8
18	48.9	43.1	76.0	70.6	91.8	88.5
19	67.7	60.6	88.2	83.9	97.3	95.0
20	17.2	14.7	60.3	54.2	87.7	85.8
$\overline{x}$	47.7	44.5	76.0	72.3	91.5	89.5
SD	18.3	20.2	10.7	13.6	5.4	6.7

Table 58. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 7, Judges = 28)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	70.5	67.5	86.3	84.4	96.5	96.5
2	53.6	52.7	78.3	77.2	93.4	94.2
3	58.0	59.6	80.5	82.5	92.6	94.6
4	71.6	78.8	89.0	92.5	97.4	98.7
5	56.5	57.4	80.2	80.4	93.8	93.4
6	40.0	40.5	66.3	68.6	87.5	89.0
7	52.5	50.0	74.5	77.8	91.6	93.3
8	60.5	60.0	78.6	79.5	92.7	93.2
9	57.7	53.2	81.2	78.9	93.3	93.3
10	60.3	60.3	83.6	83.3	96.5	96.0
11	50.9	43.6	75.3	70.0	92.1	90.0
12	36.4	26.8	63.5	54.8	85.8	80.9
13	40.9	33.9	73.8	66.8	91.9	89.1
14	68.2	57.8	87.8	80.5	97.3	94.8
15	48.2	31.4	77.9	63.9	92.4	86.1
16	36.1	33.6	65.2	62.5	85.3	84.5
17	28.9	23.0	60.2	49.8	82.5	75.5
18	35.1	30.1	70.0	66.1	91.6	88.9
19	19.5	14.0	50.7	44.3	80.3	75.2
20	42.3	33.2	70.3	62.9	89.5	86.6
21	16.2	9.8	48.9	37.2	79.9	69.3
$\bar{x}$	47.8	43.7	73.4	69.7	90.7	88.7
SD	15.6	18.2	11.1	14.3	5.3	7.8

Table 59. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 8, Judges = 32)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	72.3	77.5	87.3	90.8	98.4	99.0
2	71.6	76.9	89.3	92.8	98.2	98.6
3	69.9	72.2	87.5	89.2	97.2	97.3
4	57.8	53.3	81.5	78.8	94.9	93.7
5	62.2	61.8	83.8	83.1	96.3	95.6
6	51.3	50.2	80.6	79.0	95.9	94.6
7	55.4	54.9	85.5	85.4	97.0	96.6
8	41.9	39.4	74.7	71.4	92.3	90.7
9	44.7	37.1	75.0	68.1	90.8	88.5
10	41.1	35.2	72.0	65.6	89.0	86.8
11	29.8	25.9	65.2	60.7	87.0	84.6
12	71.8	73.9	91.2	92.4	98.6	98.2
13	54.2	45.6	81.9	76.9	96.1	92.6
14	42.4	36.5	75.3	71.3	93.5	90.8
15	38.2	29.8	69.8	59.4	90.5	84.6
16	43.5	42.6	72.8	72.3	90.7	90.7
17	36.8	32.7	71.8	68.2	92.0	89.5
18	21.8	17.5	55.9	50.5	85.3	81.1
19	22.0	13.9	52.9	41.5	82.2	70.5
20	21.0	16.3	65.0	57.1	90.8	86.4
21	11.4	8.9	46.0	38.4	79.1	75.3
$\bar{x}$	45.8	43.0	74.5	71.1	92.2	89.8
SD	18.2	21.1	12.3	15.8	5.4	7.6

Table 60. Expected Proportion-Correct Scores for the Basic, Proficient, and Advanced Levels (Grade = 12, Block = 9, Judges = 29)

	Basic		Proficient		Advanced	
Item	1st Rating	2nd Rating	1st Rating	2nd Rating	1st Rating	2nd Rating
1	60.1	59.3	79.9	79.2	94.6	93.6
2	57.9	64.7	79.1	84.8	93.6	96.8
3	40.0	33.1	67.0	61.3	84.3	81.2
4	56.0	57.2	78.4	80.9	93.5	94.2
5	64.4	66.7	83.2	84.8	96.6	96.9
6	61.7	59.4	83.1	81.1	96.3	94.1
7	24.1	16.5	52.3	41.9	77.0	67.3
8	28.3	21.6	58.3	53.9	83.8	80.1
9	25.5	15.6	61.8	54.2	84.9	79.0
10	47.1	44.5	71.1	69.1	89.9	88.1
11	23.4	15.8	50.2	45.2	75.2	70.1
12	33.0	29.0	67.7	63.4	91.1	88.9
13	45.5	37.8	73.2	66.1	90.1	87.0
14	24.7	17.3	55.2	49.5	82.3	78.2
15	25.6	17.7	52.7	44.7	79.4	71.5
16	40.9	35.6	71.3	65.5	91.8	89.6
17	28.5	16.7	52.1	40.3	75.1	64.1
18	20.2	14.5	53.9	43.5	82.6	73.2
19	19.4	16.1	42.7	39.7	74.2	72.2
20	16.8	11.4	41.4	37.3	72.7	68.9
$\bar{x}$	37.2	32.5	63.7	59.3	85.5	81.8
SD	15.9	19.4	13.4	16.5	8.0	10.8

Table 61. Summary of Grade 4 Achievement Levels at the Block Level for First and Second Ratings

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	3	CT	7	52.1	13.5	7	50.7	9.9
		MI	10	57.9	14.4	10	57.8	10.0
		CA	7	62.5	17.0	7	59.7	14.6
		FL	6	55.8	22.7	6	50.8	20.6
		Total	30	57.2	16.2	30	55.2	13.6
Proficient	3	CT	7	73.8	11.7	7	74.0	6.8
		MI	10	80.7	10.5	10	79.3	6.2
		CA	7	79.6	13.6	7	76.7	10.8
		FL	6	79.1	22.2	6	77.2	21.7
		Total	30	78.5	13.9	30	77.0	11.4
Advanced	3	CT	7	91.2	6.9	7	92.2	4.5
		MI	10	95.0	2.7	10	92.5	3.7
		CA	7	91.8	8.4	7	89.3	6.2
		FL	6	90.9	15.8	6	89.1	15.7
		Total	30	92.5	8.5	30	91.0	7.8
Basic	4	CT	9	30.7	10.6	9	27.6	9.9
		MI	5	37.5	5.5	5	33.5	5.9
		CA	7	42.5	12.5	7	40.1	13.1
		FL	4	30.2	12.3	4	21.8	7.4
		Total	25	35.3	11.3	25	31.3	11.5
Proficient	4	CT	9	56.8	8.8	9	50.4	10.5
		MI	5	70.3	7.4	5	66.5	3.7
		CA	7	66.0	13.3	7	62.5	15.4
		FL	4	65.8	19.2	4	55.3	19.0
		Total	25	63.4	12.4	25	57.8	13.7
Advanced	4	CT	9	79.9	3.9	9	72.5	12.9
		MI	5	86.8	4.1	5	84.3	3.6
		CA	7	84.8	8.4	7	81.7	10.1
		FL	4	87.3	9.6	4	79.0	5.5
		Total	25	83.8	6.9	25	78.5	10.5

Table 61. Summary of Grade 4 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	5	CT	8	30.4	10.9	8	25.3	9.8
		MI	8	45.1	8.7	8	34.0	5.1
		CA	8	48.3	18.5	8	43.5	17.5
		FL	6	58.3	22.2	6	48.9	19.5
		Total	30	44.7	17.7	30	37.2	15.8
Proficient	5	CT	8	59.2	11.3	8	50.9	12.1
		MI	8	73.1	8.5	8	61.6	9.9
		CA	8	70.4	14.4	8	64.5	14.2
		FL	6	79.8	21.4	6	72.3	19.0
		Total	30	70.0	15.2	30	61.7	15.1
Advanced	5	CT	8	88.0	4.3	8	81.3	9.0
		MI	8	88.7	6.7	8	79.2	6.6
		CA	8	88.3	6.4	8	83.3	8.5
		FL	6	90.9	14.2	6	87.0	13.6
		Total	30	88.8	7.8	30	82.4	9.4
Basic	6	CT	8	36.8	12.4	8	28.7	13.0
		MI	5	50.5	16.6	5	39.6	4.8
		CA	5	48.4	16.3	5	44.5	16.6
		FL	8	38.3	23.7	8	26.8	19.3
		Total	26	42.1	18.0	26	33.3	16.0
Proficient	6	CT	8	63.8	12.3	8	56.4	10.9
		MI	5	76.5	13.6	5	67.9	11.8
		CA	5	69.3	8.0	5	64.4	10.0
		FL	8	73.4	19.4	8	60.2	14.5
		Total	26	70.3	14.6	26	61.3	12.2
Advanced	6	CT	8	86.4	5.5	8	80.3	7.9
		MI	5	90.3	6.2	5	83.7	12.2
		CA	5	86.8	4.5	5	83.9	6.3
		FL	8	90.9	11.7	8	81.3	8.8
		Total	26	88.6	7.8	26	81.9	8.5

Table 61. Summary of Grade 4 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	7	CT	7	41.7	17.4	7	40.3	15.9
		MI	7	42.3	16.0	7	52.4	11.0
		CA	4	40.8	13.6	4	42.6	10.6
		FL	4	54.8	24.7	4	36.8	13.5
		Total	22	42.6	17.0	22	43.3	14.1
Proficient	7	CT	7	68.9	15.9	7	66.7	15.1
		MI	7	65.0	16.5	7	73.7	7.5
		CA	4	62.2	9.2	4	62.9	5.8
		FL	4	82.2	11.3	4	72.9	9.3
		Total	22	66.8	16.1	22	67.7	12.8
Advanced	7	CT	7	89.6	8.0	7	87.9	8.2
		MI	7	86.1	8.1	7	89.6	4.6
		CA	4	81.5	6.8	4	82.7	3.2
		FL	4	96.7	1.9	4	92.6	4.8
		Total	22	86.7	9.9	22	86.9	8.7
Basic	8	CT	8	45.0	14.9	8	41.9	13.9
		MI	9	54.7	18.4	9	53.1	14.1
		CA	8	54.8	18.3	8	51.6	17.4
		FL	8	50.3	23.8	8	42.8	20.3
		Total	33	50.4	19.0	33	46.8	16.9
Proficient	8	CT	8	71.1	13.2	8	64.7	11.9
		MI	9	76.7	14.2	9	76.0	9.6
		CA	8	75.2	13.3	8	73.3	13.3
		FL	8	72.8	18.7	8	64.6	13.8
		Total	33	72.8	15.8	33	68.8	14.1
Advanced	8	CT	8	89.9	6.2	8	83.7	13.0
		MI	9	92.9	5.1	9	90.7	7.4
		CA	8	90.7	6.9	8	89.7	7.0
		FL	8	91.7	10.5	8	86.9	6.8
		Total	33	90.2	9.4	33	86.8	10.7

Table 61. Summary of Grade 4 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\overline{x}$	SD	N	$\overline{x}$	SD
Basic	9	CT	7	42.3	6.3	7	33.9	9.6
		MI	10	43.1	6.8	10	37.9	8.1
		CA	6	59.6	22.1	6	56.0	20.0
		FL	6	53.1	22.2	6	39.5	16.1
		Total	29	48.4	15.8	29	41.0	15.0
Proficient	9	CT	7	66.6	7.1	7	56.9	14.5
		MI	10	72.3	7.5	10	67.0	7.2
		CA	6	79.8	16.1	6	76.1	15.8
		FL	6	75.2	13.8	6	63.5	13.6
		Total	29	73.1	11.5	29	65.7	13.6
Advanced	9	CT	7	85.2	9.2	7	74.1	17.2
		MI	10	90.9	3.5	10	86.5	6.7
		CA	6	93.5	9.1	6	90.0	9.7
		FL	6	93.7	3.9	6	86.8	7.1
		Total	29	90.6	7.1	29	84.3	11.8

Table 62. Summary of Grade 8 Achievement Levels at the Block Level for First and Second Ratings

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	3	CT	7	59.1	13.5	7	58.3	8.5
		MI	5	62.2	8.5	5	60.9	22.1
		CA	7	50.5	21.9	7	49.9	14.9
		FL	8	42.7	14.9	8	45.7	12.3
		Total	27	52.6	17.5	27	52.9	14.9
Proficient	3	CT	7	79.6	5.3	7	78.8	5.5
		MI	5	79.6	15.7	5	77.6	16.8
		CA	7	74.3	9.9	7	75.0	9.2
		FL	8	76.7	6.3	8	74.1	6.0
		Total	27	77.3	9.1	27	76.1	9.1
Advanced	3	CT	7	92.5	4.6	7	92.4	3.5
		MI	5	91.0	8.1	5	88.7	9.9
		CA	7	91.3	6.1	7	92.7	6.1
		FL	8	95.2	1.2	8	93.3	2.2
		Total	27	92.7	5.2	27	91.8	5.5
Basic	4	CT	7	48.1	14.5	7	47.5	12.8
		MI	7	42.1	18.6	7	42.6	13.6
		CA	9	40.0	13.7	9	38.1	9.8
		FL	7	44.4	18.2	7	41.7	16.2
		Total	30	43.4	15.7	30	42.2	12.8
Proficient	4	CT	7	70.7	9.4	7	69.7	8.6
		MI	7	66.2	12.9	7	67.6	9.2
		CA	9	67.1	8.0	9	63.4	6.6
		FL	7	72.9	16.6	9	68.1	17.7
		Total	30	69.1	11.6	30	57.0	10.7
Advanced	4	CT	7	85.9	5.5	7	85.8	5.0
		MI	7	85.4	6.2	7	85.8	7.2
		CA	9	85.7	5.6	9	83.1	6.0
		FL	7	95.9	2.9	7	92.7	4.4
		Total	30	88.0	6.6	30	86.6	6.6

Table 62. Summary of Grade 8 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	5	CT	7	47.5	14.5	7	44.0	14.7
		MI	7	53.3	16.2	7	50.8	18.4
		CA	9	51.7	12.8	9	46.4	11.9
		FL	7	60.8	10.0	9	45.5	8.9
		Total	30	53.2	13.6	30	46.6	13.3
Proficient	5	CT	7	70.5	8.5	7	68.6	10.3
		MI	7	78.7	8.9	7	75.7	10.7
		CA	9	77.4	7.0	9	74.2	5.7
		FL	7	80.4	8.1	7	68.5	3.2
		Total	30	76.8	8.5	30	71.9	8.2
Advanced	5	CT	7	89.0	4.7	7	88.1	5.5
		MI	7	93.0	5.2	7	91.7	5.6
		CA	9	94.0	2.8	9	91.1	2.3
		FL	7	92.8	5.8	7	86.7	2.66
		Total	30	92.3	4.8	30	89.5	4.5
Basic	6	CT	6	64.6	14.9	6	66.2	12.7
		MI	9	40.0	15.0	9	50.0	13.5
		CA	6	49.4	12.0	6	51.5	10.4
		FL	7	51.8	12.3	7	47.1	13.3
		Total	28	53.3	14.2	28	51.8	12.9
Proficient	6	CT	6	85.2	6.2	6	81.9	4.4
		MI	9	74.7	9.1	9	72.8	9.6
		CA	6	71.6	8.8	6	72.9	8.0
		FL	7	76.0	7.0	7	69.2	9.3
		Total	28	76.6	9.0	28	73.9	9.1
Advanced	6	CT	6	95.6	2.2	6	95.0	2.2
		MI	9	90.2	5.2	9	89.1	5.9
		CA	6	89.2	4.3	6	89.5	5.7
		FL	7	85.2	2.5	7	92.4	4.2
		Total	28	92.4	4.7	28	91.3	5.2

Table 62. Summary of Grade 8 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	7	CT	8	45.9	12.1	8	39.9	11.6
		MI	5	42.7	9.3	5	41.2	7.2
		CA	8	43.6	13.9	8	41.2	10.3
		FL	6	39.1	15.3	8	29.7	11.2
		Total	27	41.1	12.5	27	38.3	10.9
Proficient	7	CT	8	67.6	8.7	8	61.9	7.3
		MI	5	69.8	13.2	5	69.7	12.5
		CA	8	73.2	8.8	8	70.9	8.0
		FL	6	74.0	4.5	6	57.0	9.6
		Total	27	71.1	8.9	27	64.9	10.3
Advanced	7	CT	8	88.2	4.3	8	84.9	3.3
		MI	5	88.7	7.0	5	88.1	4.9
		CA	8	93.1	2.3	8	83.1	8.2
		FL	6	93.2	3.4	6	83.1	8.2
		Total	27	90.9	4.7	27	87.0	5.7
Basic	8	CT	7	50.2	13.0	7	46.3	9.9
		MI	5	48.0	9.5	5	44.2	11.9
		CA	7	46.2	7.8	7	44.1	6.9
		FL	6	44.8	21.0	6	44.4	16.5
		Total	25	47.3	13.0	25	44.8	10.9
Proficient	8	CT	7	73.5	6.6	7	70.6	6.6
		MI	5	73.5	9.1	5	68.8	13.1
		CA	7	69.1	6.0	7	65.8	7.0
		FL	6	76.8	13.6	6	71.4	15.2
		Total	25	73.1	9.0	25	69.0	10.3
Advanced	8	CT	7	89.3	6.4	7	88.6	6.4
		MI	5	91.4	8.0	5	87.3	7.5
		CA	7	87.3	4.7	7	83.9	6.2
		FL	6	91.5	10.3	6	85.2	12.2
		Total	25	89.7	7.2	25	86.2	8.0

Table 62. Summary of Grade 8 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	9	CT	6	52.8	16.6	6	47.4	15.7
		MI	7	50.3	19.1	7	51.8	16.1
		CA	8	33.8	12.4	8	32.0	9.9
		FL	7	48.2	7.3	7	42.0	8.5
		Total	28	45.6	15.6	28	42.8	14.3
Proficient	9	CT	6	77.0	9.9	6	70.1	10.6
		MI	7	77.4	13.1	7	78.5	8.5
		CA	8	63.3	14.1	8	62.0	11.5
		FL	7	77.3	6.8	7	70.5	10.4
		Total	28	73.3	12.6	28	70.0	11.5
Advanced	9	CT	6	91.9	4.6	6	88.3	5.5
		MI	7	92.7	3.4	7	93.8	3.0
		CA	8	86.3	10.6	8	85.8	7.0
		FL	7	94.3	3.5	7	91.7	5.7
		Total	28	91.1	7.0	28	89.8	6.28

Table 63. Summary of Grade 12 Achievement Levels at the Block Level for First and Second Ratings

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	3	CT	9	47.9	13.5	9	47.4	13.5
		MI	6	55.1	18.2	6	55.1	11.8
		CA	9	41.2	15.4	9	46.8	10.7
		FL	8	50.3	12.7	8	45.5	8.4
		Total	32	48.0	14.9	32	48.2	11.3
Proficient	3	CT	9	78.1	8.7	9	75.2	11.1
		MI	6	80.7	11.5	6	75.5	8.2
		CA	9	72.6	15.0	9	73.5	9.0
		FL	8	81.3	9.2	8	77.5	8.5
		Total	32	77.8	11.4	32	76.1	9.2
Advanced	3	CT	9	93.2	4.0	9	91.3	5.2
		MI	6	95.3	4.6	6	94.6	3.6
		CA	9	94.8	4.3	9	94.1	4.4
		FL	8	94.6	4.1	8	93.4	3.9
		Total	32	94.4	4.1	32	93.3	4.4
Basic	4	CT	9	53.6	12.6	9	53.2	9.9
		MI	7	58.1	18.5	7	57.6	17.4
		CA	9	51.2	5.9	9	53.3	6.2
		FL	6	47.1	17.3	6	45.3	15.0
		Total	31	52.7	13.8	31	52.7	12.3
Proficient	4	CT	9	76.8	11.8	9	74.7	9.5
		MI	7	81.5	14.0	7	80.0	14.6
		CA	9	75.8	8.3	9	75.9	8.9
		FL	6	73.7	10.6	6	71.8	11.3
		Total	31	77.0	11.0	31	75.7	10.8
Advanced	4	CT	9	89.4	10.1	9	88.0	8.3
		MI	7	94.7	5.8	7	94.0	7.0
		CA	9	92.3	5.6	9	90.9	6.5
		FL	6	93.5	3.4	6	92.6	3.9
		Total	31	92.2	7.0	31	91.1	6.9

Table 63. Summary of Grade 12 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	5	CT	9	49.2	15.0	9	42.1	13.2
		MI	4	51.9	9.1	4	50.1	5.8
		CA	9	53.7	18.7	9	49.1	19.3
		FL	7	37.3	17.1	7	34.9	17.0
		Total	29	48.1	16.6	29	43.6	16.0
Proficient	5	CT	9	75.2	9.6	9	68.4	13.3
		MI	4	74.9	5.8	4	74.5	4.7
		CA	9	74.5	20.7	9	69.0	20.0
		FL	7	70.9	10.6	7	67.6	11.7
		Total	29	73.9	13.4	29	69.3	14.2
Advanced	5	CT	9	91.0	4.9	9	85.1	8.9
		MI	4	93.0	2.3	4	92.2	2.2
		CA	9	92.5	6.1	9	87.5	9.9
		FL	7	90.2	7.6	7	87.7	8.9
		Total	29	91.6	5.6	29	87.4	8.6
Basic	6	CT	8	41.9	9.6	8	38.6	7.4
		MI	7	42.8	12.5	7	42.0	10.9
		CA	8	57.6	13.6	8	56.5	14.1
		FL	6	47.8	16.0	6	39.2	14.3
		Total	28	47.7	13.8	29	44.5	13.6
Proficient	6	CT	8	68.6	8.9	8	63.4	7.4
		MI	7	72.8	10.6	7	70.4	9.1
		CA	8	85.3	7.5	8	84.0	9.2
		FL	6	77.1	9.0	6	70.7	10.5
		Total	29	76.0	10.8	29	72.3	11.7
Advanced	6	CT	8	85.2	7.5	8	82.0	6.1
		MI	7	91.1	5.6	7	89.8	5.9
		CA	8	96.9	2.3	8	96.4	2.9
		FL	6	93.3	4.6	6	90.0	6.9
		Total	29	91.5	6.8	29	89.5	7.6

Table 63. Summary of Grade 12 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	7	CT	8	47.1	10.6	8	43.7	7.0
		MI	5	47.1	10.5	5	44.9	9.0
		CA	8	51.4	25.7	8	51.3	22.7
		FL	7	45.0	22.7	7	34.1	15.1
		Total	28	47.8	18.4	28	43.7	15.8
Proficient	7	CT	8	71.6	13.1	8	68.4	10.8
		MI	5	72.1	13.4	5	70.1	9.6
		CA	8	72.5	17.3	8	72.2	13.9
		FL	7	77.4	7.8	7	68.0	7.6
		Total	28	73.4	12.9	28	69.7	10.5
Advanced	7	CT	8	88.2	8.3	8	86.6	8.2
		MI	5	91.3	8.2	5	89.8	6.6
		CA	8	90.1	9.4	8	89.4	7.9
		FL	7	93.7	4.3	7	89.6	5.9
		Total	28	90.7	7.7	28	88.7	7.1
Basic	8	CT	9	40.3	11.9	9	41.0	7.2
		MI	7	54.3	19.3	7	52.4	16.7
		CA	9	41.1	22.7	9	41.4	20.2
		FL	7	43.7	16.9	7	38.0	11.9
		Total	32	45.7	17.9	32	43.0	15.1
Proficient	8	CT	9	72.2	8.6	9	67.9	6.9
		MI	7	80.2	12.6	7	78.7	12.7
		CA	9	69.5	16.1	9	68.6	15.9
		FL	7	78.2	13.1	7	70.7	11.5
		Total	32	74.5	13.0	32	71.1	12.3
Advanced	8	CT	9	87.9	5.4	9	84.5	6.4
		MI	7	94.2	5.8	7	93.3	5.9
		CA	9	92.0	5.6	9	90.0	7.4
		FL	7	95.9	3.6	7	92.8	3.4
		Total	32	92.2	5.8	32	89.8	6.8

Table 63. Summary of Grade 12 Achievement Levels at the Block Level for First and Second Ratings--Continued

			1st Rating			2nd Rating		
Level	Block	Site	N	$\bar{x}$	SD	N	$\bar{x}$	SD
Basic	9	CT	8	30.7	12.3	8	30.0	11.2
		MI	6	48.9	21.2	6	42.2	19.5
		CA	8	36.9	21.2	8	34.2	23.0
		FL	7	34.7	13.8	7	25.1	7.3
		Total	29	37.2	17.7	29	32.5	16.7
Proficient	9	CT	8	57.5	14.0	8	55.8	13.8
		MI	6	74.5	18.2	6	70.7	18.4
		CA	8	62.1	18.5	8	57.5	21.0
		FL	7	63.5	11.5	7	55.7	9.11
		Total	29	63.7	16.1	29	59.3	16.5
Advanced	9	CT	8	79.4	8.5	8	77.1	8.2
		MI	6	90.0	8.6	6	89.0	9.3
		CA	8	87.4	7.9	8	82.0	14.3
		FL	7	86.2	5.8	7	80.6	6.9
		Total	29	85.4	8.4	29	81.7	10.5

Table 64. Summary of Final Achievement Levels

			Basic			Proficient			Advanced		
Grade	Site	N	$\overline{x}$	P_{50}	SD	$\overline{x}$	P_{50}	SD	$\overline{x}$	P_{50}	SD
4	CT	18	38.1	40.0	9.5	64.1	65.0	8.9	85.6	85.0	5.1
	MI	17	49.8	49.0	6.9	74.0	72.0	6.1	88.2	88.0	3.7
	CA	16	53.4	51.0	10.9	72.2	71.5	9.0	88.1	89.0	5.1
	FL	14	44.0	41.0	13.9	67.7	70.0	13.9	86.4	90.0	9.3
	Total	65	45.0	44.0	12.1	68.0	70.0	10.3	86.7	86.0	6.3
8	CT	16	51.7	50.0	3.8	73.4	72.0	3.2	89.1	90.0	2.9
	MI	20	51.7	50.0	9.7	75.2	78.0	6.3	88.5	90.0	8.9
	CA	20	44.8	42.5	7.1	70.4	70.0	5.6	88.4	89.0	3.7
	FL	17	45.5	45.0	6.6	71.4	69.0	6.0	91.2	92.0	3.0
	Total	73	48.0	48.0	7.7	72.1	72.0	5.6	89.0	90.0	5.5
12	CT	20	46.1	45.0	5.2	71.1	72.0	6.1	87.0	87.5	4.7
	MI	17	48.6	47.0	8.2	73.6	72.0	6.8	90.1	90.0	3.2
	CA	21	51.3	50.0	14.0	74.2	75.0	12.5	91.3	90.0	6.1
	FL	15	38.3	39.0	8.9	70.3	70.0	5.9	89.4	90.0	5.3
	Total	73	46.6	45.0	10.8	72.6	72.0	8.6	88.4	90.0	5.3

Table 65. Summary of Confidence Levels on the Final Ratings

				Level of Confidence					
Grade	Level	Site	N	1	2	3	4	$\bar{x}$	SD
4	Basic	CT	18	6	39	39	17	2.7	0.8
		MI	17	0	6	65	29	3.2	0.6
		CA	16	0	44	50	6	2.6	0.6
		FL	14	0	21	36	43	3.2	0.8
		Total	65	2	33	44	22	2.9	0.8
	Proficient	CT	18	0	28	61	11	3.2	0.6
		MI	17	0	12	35	53	3.4	0.7
		CA	16	0	31	38	31	3.0	0.8
		FL	14	0	7	57	36	3.3	0.6
		Total	65	0	24	46	31	3.1	0.7
	Advanced	CT	18	0	28	22	50	3.2	0.9
		MI	17	6	12	35	47	3.2	0.9
		CA	16	0	25	31	44	3.2	0.8
		FL	14	0	7	50	43	3.4	0.6
		Total	65	2	20	31	4	3.2	0.8
8	Basic	CT	16	13	25	44	19	2.7	0.9
		MI	20	15	30	45	10	2.5	0.9
		CA	20	5	50	35	10	2.5	0.8
		FL	17	6	24	47	24	2.9	0.9
		Total	73	10	31	44	15	2.6	0.9
	Proficient	CT	16	0	13	69	19	3.1	0.6
		MI	20	5	20	65	10	2.8	0.7
		CA	20	0	20	60	20	3.0	0.6
		FL	17	0	29	53	18	2.9	0.7
		Total	73	2	21	63	15	2.9	0.6
	Advanced	CT	16	0	6	25	69	3.6	0.6
		MI	20	5	10	40	45	3.3	0.9
		CA	20	0	5	25	70	3.7	0.6
		FL	17	0	18	24	59	3.4	0.8
		Total	73	2	10	27	62	3.5	0.7

Table 65. Summary of Confidence Levels on the Final Ratings--Continued

				Level of Confidence					
Grade	Level	Site	N	1	2	3	4	$\overline{x}$	SD
12	Basic	CT	20	0	32	42	26	2.9	0.8
		MI	17	0	35	41	24	2.9	0.8
		CA	21	0	24	48	29	3.0	0.7
		FL	15	7	13	60	20	2.9	0.8
		Total	73	1	27	46	26	3.0	0.8
	Proficient	CT	20	0	16	63	21	3.1	0.6
		MI	17	6	6	47	41	3.2	0.8
		CA	21	0	5	62	33	3.3	0.6
		FL	15	0	0	67	33	3.3	0.5
		Total	73	1	7	61	30	3.2	0.6
	Advanced	CT	20	0	11	42	47	3.4	0.7
		MI	17	0	6	29	65	3.6	0.6
		CA	21	0	0	33	67	3.7	0.5
		FL	15	0	0	71	29	3.3	0.5
		Total	73	0	4	44	52	3.5	0.6

Table 66. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process

Question		Site				
		Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
1.	What is your overall impression of the training you received today for setting achievement levels?					
	a. appropriate	83	69	70	83	76
	b. somewhat appropriate	17	29	29	17	23
	c. not appropriate	0	2	0	0	1
2.	How clear were you about NAGB's definition of the Basic student?					
	a. not at all clear	2	7	7	0	4
	b. somewhat clear	26	42	41	32	35
	c. clear	52	44	38	45	44
	d. very clear	20	7	14	23	16
3.	How clear were you about NAGB's definition of the Proficient student?					
	a. not at all clear	0	4	0	0	1
	b. somewhat clear	13	27	30	28	25
	c. clear	65	55	45	45	52
	d. very clear	22	15	25	28	22
4.	How clear were you about NAGB's definition of the Advanced student?					
	a. not at all clear	0	2	0	0	1
	b. somewhat clear	17	18	18	23	19
	c. clear	50	58	46	34	48
	d. very clear	33	22	36	43	33

Table 66. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process--Continued

	Site				
Question	Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
5. How would you judge the time allotted today to set achievement levels?					
a. not enough time	6	9	18	13	11
b. too much time	2	7	2	2	3
c. about the right amount of time	93	84	78	85	83
6. How would you judge your level of understanding of the achievement level setting process implemented today?					
a. low	4	0	2	0	1
b. medium	37	53	30	4	40
c. high	59	47	66	60	58
7. Which factors influenced the achievement levels that you set today? (Circle all choices which apply.)					
a. the definitions of basic, proficient, and advanced students	91	87	93	85	89
b. the content of the items	85	78	89	77	83
c. my perception of the difficulty of items	87	93	96	92	92
d. actual student performance on the items	82	73	70	72	74
e. persons working with the same test booklet	44	33	48	36	41
f. persons working at the same grade level as myself	44	47	52	36	45
g. persons working at the other grade levels	9	9	11	6	9
h. other (Please specify: ____________________)	15	9	14	6	11

Table 66. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process--Continued

	Site				
Question	Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
9. Do you believe that achievement levels will be useful in interpreting student performance on the 1990 NAEP Mathematics Assessment?					
a. Definitely Yes	41	26	43	36	36
b. Probably Yes	57	51	41	55	51
c. Unsure	2	20	16	6	11
d. Probably No	0	4	0	2	1
e. Definitely No	0	0	0	0	0
10. How successful do you believe the process was today in setting achievement levels?					
a. very successful	20	18	16	26	20
b. successful	65	49	59	53	57
c. somewhat successful	15	31	25	21	23
d. not successful at all	0	2	0	0	1

Table 67. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process

Question	Site				
	Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
13. Which best describes you?					
a. White	83	89	66	66	77
b. Black	11	11	13	28	15
c. Hispanic	2	0	9	4	4
d. Asian	2	0	7	2	2
e. Native American	0	0	0	0	0
f. Other: ________________	0	0	4	0	1
14. What is your gender?					
a. Male	43	38	36	40	40
b. Female	57	62	63	60	60
15. Which type of organization do you represent here today?					
a. business	4	7	2	9	5
b. industry	0	4	0	2	1
c. school board	0	0	9	2	3
d. parents	0	4	2	2	2
e. educators	28	27	21	19	24
f. math educators	67	58	64	62	63
g. other: ________________	0	0	2	4	1
16. Which best describes your current professional status?					
a. Mathematics teacher in grade 4, 8, or 12	57	69	68	83	69
b. Mathematics supervisor, elementary	6	4	4	0	3
c. Mathematics supervisor, secondary	7	4	2	0	3
d. Mathematics supervisor, K-12	7	0	0	0	2
e. School administrator	6	0	0	0	1
f. Non-educator	4	6	11	11	8
g. Other: ________________	11	18	16	6	13

Table 67. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process--Continued

Question	Site				
	Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
17. What type of community do you work/teach in?					
a. urban or mostly urban	37	33	52	49	43
b. suburban	39	47	43	38	42
c. rural or mostly rural	22	20	5	13	15
18. How large is the community in which you work/teach?					
a. small town	33	36	7	9	22
b. large town	32	19	13	23	22
c. medium city	19	40	39	30	32
d. large city	15	6	41	38	25
19. Approximately how many students do you teach? __________					
20. What ability levels do you mostly teach?					
a. average mainstream students	59	50	50	34	49
b. below average mainstream students	15	21	24	16	19
c. above average mainstream students	27	29	20	34	27
d. special needs students	0	0	7	16	5
21. How long have you been teaching?					
a. 1 to 3 years	2	2	0	10	4
b. 4 to 10 years	27	15	28	23	24
c. 11 to 20 years	33	38	45	39	37
d. 21 years or more	38	45	26	28	35

Table 67. Summary of Participant Evaluations of the NAGB Achievement Level Setting Process--Continued

Question	Site				
	Connecticut (N=54)	Michigan (N=55)	California (N=56)	Florida (N=47)	Total (N=212)
22. Which best describes the organization for whom you currently work?					
a. non-profit organization	0	0	38	44	24
b. branch of the military	0	0	0	0	0
c. federal, state, local government	40	14	0	22	17
d. large corporation	0	71	0	33	28
e. small business (less than 100 employees)	20	0	13	0	7
f. self-employed	0	14	13	0	7
g. other: ______________________	40	0	38	0	17

Table 68. Summary of the Achievement Level Review Results (Grade 4, N=66)

		Percent of Responses*		
Level	Skill	Yes	No	Unsure
Basic	1	100	0	0
	2	96	0	4
	3	100	0	0
	4	93	4	4
	5	96	0	4
	6	96	0	4
	7	93	2	5
	8	98	0	2
	9	100	0	0
	10	96	0	4
Proficient	11	98	0	2
	12	96	0	4
	13	95	0	5
	14	91	0	9
	15	86	2	13
	16	98	0	2
	17	98	0	2
	18	88	2	11
	19	96	0	4
	20	100	0	0
	21	100	0	0
	22	96	0	4
	23	100	0	0
	24	94	0	4
	25	100	0	0
	26	98	0	2
	27	86	0	14
	28	86	0	14
	29	100	0	0
Advanced	30	100	0	0
	31	96	0	4
	32	91	0	9
	33	91	2	7
	34	96	0	4
	35	76	2	22
	36	84	0	16
	37	95	0	6

*The question was: Should this skill be included in the definition?

Table 69. Summary of the Achievement Level Review Results (Grade 8, N=72)

		Percent of Responses*		
Level	Skill	Yes	No	Unsure
Basic	1	97	0	3
	2	90	6	5
	3	93	0	8
	4	97	0	3
	5	84	5	12
	6	93	0	8
	7	84	5	12
	8	92	2	6
	9	83	3	14
	10	94	2	5
	11	91	2	8
	12	100	0	0
Proficient	13	99	0	2
	14	71	12	17
	15	96	3	2
	16	88	8	5
	17	76	10	13
	18	100	0	0
	19	69	13	18
	20	77	6	18
	21	79	6	15
	22	97	0	3
	23	100	0	0
	24	87	6	8
	25	100	0	0
Advanced	26	92	0	8
	27	99	0	2
	28	73	10	16
	29	100	0	0
	30	87	8	6
	31	79	8	13
	32	94	0	6
	33	100	0	0

*The question was: Should this skill be included in the definition?

Table 70. Summary of the Achievement Level Review Results (Grade 12, N=73)

Level	Skill	Percent of Responses* Yes	No	Unsure
Basic	1	100	0	0
	2	100	0	0
	3	99	0	1
	4	86	3	11
	5	89	6	6
	6	90	3	7
	7	99	1	0
Proficient	8	100	0	0
	9	89	3	9
	10	99	0	1
	11	93	1	6
	12	100	0	0
	13	90	3	7
	14	90	3	7
Advanced	15	100	0	0
	16	97	0	3
	17	97	1	1
	18	86	4	10
	19	99	0	1
	20	96	0	4

*The question was: Should this skill be included in the definition?

Table 71. Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 4)

Level	Block	Items	Correlation			Estimated p		1st Ratings		2nd Ratings	
			r_{E1}	r_{E2}	r_{12}	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Basic	3	19	0.68	0.89	0.92	0.60	0.21	58.4	9.9	55.7	13.0
	4	14	0.86	0.96	0.96	0.41	0.21	35.3	14.1	31.4	17.0
	5	11	0.71	0.92	0.91	0.32	0.20	45.0	9.4	38.5	12.0
	6	17	0.46	0.88	0.78	0.37	0.21	40.4	7.6	32.8	9.8
	7	18	0.66	0.93	0.84	0.58	0.17	44.5	8.5	42.6	11.7
	8	15	0.84	0.93	0.93	0.53	0.19	50.1	7.3	45.8	11.1
	9	15	0.56	0.86	0.88	0.43	0.24	49.2	10.4	42.4	13.1
Proficient	3	19	0.66	0.90	0.91	0.60	0.21	80.0	6.4	78.4	8.4
	4	14	0.78	0.93	0.95	0.41	0.21	63.5	11.7	57.8	14.9
	5	11	0.75	0.92	0.93	0.32	0.20	69.1	8.0	61.5	11.6
	6	17	0.44	0.89	0.75	0.37	0.21	68.9	6.1	60.7	8.4
	7	18	0.58	0.89	0.86	0.58	0.17	69.5	7.1	68.4	9.0
	8	15	0.83	0.98	0.90	0.53	0.19	73.5	5.1	68.6	8.9
	9	15	0.58	0.87	0.89	0.43	0.24	73.0	8.5	66.3	12.5
Advanced	3	19	0.62	0.91	0.84	0.60	0.21	93.3	3.3	92.3	4.4
	4	14	0.66	0.89	0.91	0.41	0.21	83.8	8.2	78.5	10.4
	5	11	0.71	0.85	0.93	0.32	0.20	88.9	5.7	83.1	9.0
	6	17	0.45	0.84	0.79	0.37	0.21	88.1	4.6	82.2	6.3
	7	18	0.47	0.82	0.85	0.58	0.17	88.7	5.1	88.2	6.3
	8	15	0.89	0.95	0.88	0.53	0.19	91.1	2.8	87.5	5.1
	9	15	0.58	0.87	0.85	0.43	0.24	90.5	4.6	84.8	9.2

Table 72. Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 8)

			Correlation			Estimated p		1st Ratings		2nd Ratings	
Level	Block	Items	r_{E1}	r_{E2}	r_{12}	$\overline{x}$	SD	$\overline{x}$	SD	$\overline{x}$	SD
Basic	3	23	0.65	0.89	0.91	0.65	0.17	52.6	11.4	52.9	15.2
	4	21	0.91	0.95	0.99	0.52	0.25	43.5	16.5	42.2	18.9
	5	16	0.82	0.91	0.98	0.49	0.18	53.3	13.2	46.8	14.2
	6	21	0.67	0.88	0.71	0.61	0.21	53.3	8.4	51.8	11.4
	7	18	0.72	0.90	0.94	0.40	0.20	43.1	10.3	38.1	13.1
	8	18	0.89	0.97	0.96	0.45	0.27	47.3	14.2	44.8	18.5
	9	20	0.74	0.91	0.94	0.46	0.24	45.6	10.9	42.8	15.0
Proficient	3	23	0.63	0.91	0.88	0.65	0.17	77.3	7.2	76.2	8.8
	4	21	0.91	0.97	0.98	0.52	0.25	69.2	11.3	67.1	14.0
	5	16	0.86	0.92	0.99	0.49	0.18	77.0	8.7	72.2	10.8
	6	21	0.71	0.92	0.91	0.61	0.21	76.6	6.0	73.9	9.5
	7	18	0.70	0.85	0.95	0.40	0.20	71.4	8.5	65.4	10.7
	8	18	0.88	0.96	0.97	0.45	0.27	73.1	9.9	69.0	14.0
	9	20	0.68	0.91	0.92	0.46	0.24	73.3	6.9	70.0	10.0
Advanced	3	23	0.50	0.85	0.86	0.65	0.17	92.7	3.3	91.8	4.4
	4	21	0.89	0.96	0.98	0.52	0.25	88.2	6.0	86.7	7.8
	5	16	0.83	0.93	0.97	0.49	0.18	92.4	4.6	89.6	6.0
	6	21	0.74	0.93	0.91	0.61	0.21	92.4	3.2	91.3	5.1
	7	18	0.64	0.78	0.95	0.40	0.20	90.0	5.1	87.1	6.5
	8	18	0.84	0.93	0.97	0.45	0.27	89.7	6.1	86.2	9.3
	9	20	0.57	0.83	0.89	0.46	0.24	91.1	3.7	89.8	5.0

Table 73. Correlations Among Actual Item p-values and First and Second Ratings of Expected P-values (Grade 12)

			Correlation			Estimated p		1st Ratings		2nd Ratings	
Level	Block	Items	r_{E1}	r_{E2}	r_{12}	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
Basic	3	23	0.81	0.95	0.95	0.65	0.20	48.0	15.3	48.2	18.3
	4	22	0.84	0.92	0.98	0.64	0.25	52.7	19.8	52.7	22.5
	5	17	0.87	0.95	0.98	0.50	0.27	48.1	16.4	43.6	19.0
	6	20	0.87	0.92	0.99	0.52	0.23	47.7	18.3	44.5	20.2
	7	21	0.83	0.94	0.96	0.51	0.22	47.8	15.6	43.7	18.2
	8	21	0.93	0.97	0.99	0.48	0.28	45.8	18.2	43.0	21.1
	9	20	0.91	0.96	0.99	0.37	0.26	37.2	15.9	32.5	19.4
Proficient	3	23	0.85	0.96	0.96	0.65	0.20	77.8	10.0	76.1	12.6
	4	22	0.81	0.90	0.98	0.64	0.25	77.0	13.8	75.7	16.1
	5	17	0.91	0.98	0.97	0.50	0.27	73.9	10.7	69.2	15.0
	6	20	0.89	0.94	0.99	0.52	0.23	76.0	10.7	72.3	13.6
	7	21	0.78	0.93	0.95	0.51	0.22	73.4	11.1	69.7	14.3
	8	21	0.88	0.94	0.99	0.48	0.28	74.5	12.3	71.1	15.8
	9	20	0.87	0.95	0.98	0.37	0.26	63.7	13.4	59.3	16.5
Advanced	3	23	0.86	0.95	0.96	0.65	0.20	94.4	3.8	93.3	5.6
	4	22	0.83	0.91	0.97	0.64	0.25	92.2	6.2	91.1	8.0
	5	17	0.91	0.97	0.97	0.50	0.27	91.6	5.8	87.5	9.4
	6	20	0.88	0.94	0.99	0.52	0.23	91.5	5.4	89.5	6.7
	7	21	0.76	0.89	0.95	0.51	0.22	90.7	5.3	88.7	7.8
	8	21	0.86	0.91	0.97	0.48	0.28	92.2	5.4	89.8	7.6
	9	20	0.86	0.93	0.97	0.37	0.26	85.5	8.0	81.8	10.5

Table 74. Analysis of Final Achievement Levels for Educators and Non-Educators

		Educators				Non-Educators			
Grade	Level	N	$\overline{x}$	P_{50}	SD	N	$\overline{x}$	P_{50}	SD
4	Basic	44	45.6	45.0	12.3	3	49.3	43.0	13.7
	Proficient	44	68.8	70.0	11.1	3	70.3	69.0	9.1
	Advanced	44	87.8	86.0	6.7	3	87.3	89.0	3.8
8	Basic	59	48.0	50.0	6.9	6	50.1	51.0	4.6
	Proficient	59	72.8	73.0	5.9	6	70.8	70.5	2.6
	Advanced	59	89.3	90.0	5.9	6	89.0	88.5	2.2
12	Basic	62	46.0	45.0	10.1	7	50.3	53.0	16.0
	Proficient	62	72.2	72.0	8.1	7	73.1	73.0	12.5
	Advanced	62	89.3	90.0	4.9	7	88.7	90.0	7.9

Table 75. Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 4, 8, and 12 NAEP Test Booklets

	Placement*			Actual p-Value			Judges' Item Ratings								
	Grade			Grade			Basic			Proficient			Advanced		
Common Item	4	8	12	4	8	12	4	8	12	4	8	12	4	8	12
1	4,1	4,1	4,1	.87	.92	.93	.69	.81	.89	.88	.92	.96	.97	.98	.99
2	4,2	4,2	4,2	.76	.86	.90	.62	.78	.86	.83	.89	.95	.94	.97	.98
3	4,3	4,3	4,3	.69	.79	.86	.48	.56	.66	.71	.77	.82	.88	.90	.96
4	4,4	4,4	4,4	.44	.73	.88	.37	.61	.75	.67	.81	.89	.85	.95	.98
5	4,5	4,5	4,5	.42	.68	.81	.27	.47	.63	.55	.72	.83	.80	.91	.95
6	4,6	4,6	4,6	.31	.74	.88	.28	.56	.73	.58	.79	.89	.78	.94	.98
7	4,7	4,7	4,7	.34	.55	.69	.21	.38	.49	.44	.65	.73	.68	.85	.89
8	4,8	4,8	4,8	.33	.60	.71	.25	.50	.61	.54	.74	.83	.77	.91	.96
9	4,9	4,9	4,9	.25	.68	.82	.30	.58	.74	.57	.77	.89	.79	.95	.97
10	4,10	4,10	4,10	.30	.63	.78	.20	.43	.65	.52	.72	.83	.74	.89	.95
11	4,11	4,11	4,11	.38	.78	.88	.38	.56	.71	.59	.80	.88	.80	.94	.97
12	4,12	4,12	4,12	.24	.46	.64	.14	.34	.50	.38	.63	.76	.60	.85	.92
13	4,13	4,13	4,13	.22	.33	.47	.15	.32	.46	.42	.62	.74	.68	.85	.90
14	4,14	4,14	4,14	.19	.53	.76	.16	.41	.58	.43	.71	.84	.68	.90	.95
15	5,6	5,6	5,6	.48	.85	.86	.41	.68	.70	.67	.86	.86	.89	.96	.97
16	5,7	5,7	5,7	.21	.58	.82	.37	.53	.70	.58	.77	.88	.84	.94	.97
17	5,8	5,8	5,8	.30	.47	.59	.33	.43	.51	.58	.69	.74	.81	.88	.89
18	5,9	5,9	5,9	.23	.58	.77	.34	.52	.65	.58	.78	.85	.82	.93	.97
19	6,1	6,1	6,1	.60	.84	.92	.42	.69	.82	.67	.87	.82	.90	.97	.99
20	6,2	6,2	6,2	.24	.74	.89	.28	.61	.76	.54	.82	.91	.78	.94	.98
Means:				.39	.67	.79	.33	.54	.67	.59	.77	.85	.80	.92	.96

*Block, Item Number in the Block

NOTE: 20 common items.

Table 76. Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 4 and 8 NAEP Test Booklets

	Placement		Actual p-Value		Judges' Item Ratings					
	Grade		Grade		Basic		Proficient		Advanced	
Common Item	4	8	4	8	4	8	4	8	4	8
1	5,1	5,1	.49	.75	.51	.68	.70	.87	.90	.98
2	5,2	5,2	.05	.63	.22	.58	.43	.81	.66	.96
3	5,3	5,3	.23	.43	.32	.44	.54	.70	.78	.88
4	5,4	5,4	.48	.52	.51	.64	.75	.85	.92	.96
5	5,5	5,5	.15	.60	.37	.64	.62	.86	.86	.96
6	6,3	6,3	.87	.95	.59	.74	.81	.88	.93	.97
7	6,4	6,4	.07	.12	.29	.41	.56	.63	.78	.81
8	6,5	6,5	.65	.90	.40	.70	.68	.87	.86	.98
9	6,6	6,6	.48	.78	.35	.58	.65	.81	.85	.95
10	6,7	6,7	.16	.55	.31	.53	.59	.76	.82	.93
11	6,8	6,8	.43	.53	.33	.44	.61	.68	.84	.90
12	6,9	6,9	.46	.75	.35	.55	.65	.78	.87	.94
13	6,10	6,1	.41	.68	.42	.59	.67	.80	.88	.96
14	6,11	6,1	.51	.75	.41	.58	.69	.79	.87	.95
15	6,12	6,1	.32	.69	.27	.53	.58	.77	.82	.93
16	6,13	6,13	.28	.65	.27	.52	.56	.75	.79	.92
17	6,14	6,14	.19	.52	.19	.40	.46	.63	.69	.85
18	6,15	6,15	.25	.61	.23	.45	.52	.70	.76	.90
19	6,16	6,16	.21	.70	.23	.53	.53	.77	.75	.95
20	6,17	6,17	.15	.58	.24	.50	.54	.74	.78	.93
Means:			.34	.63	.34	.55	.61	.77	.82	.93

NOTE: 20 common items.

Table 77. Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 8 and 12 NAEP Test Booklets

	Placement		Actual p-Value		Judges' Item Ratings					
	Grade		Grade		Basic		Proficient		Advanced	
Common Item	8	12	8	12	8	12	8	12	8	12
1	7,1	7,1	.58	.72	.51	.68	.77	.84	.95	.97
2	7,2	7,2	.48	.63	.39	.53	.71	.77	.92	.94
3	7,3	7,3	.58	.76	.46	.60	.74	.83	.91	.95
4	7,4	7,4	.92	.96	.73	.79	.89	.93	.98	.98
5	7,5	7,5	.44	.69	.45	.57	.71	.80	.92	.93
6	7,6	7,6	.43	.60	.34	.41	.61	.69	.86	.89
7	7,7	7,7	.43	.65	.39	.50	.67	.78	.90	.93
8	7,8	7,8	.55	.70	.41	.60	.70	.80	.92	.93
9	7,9	7,9	.41	.55	.38	.53	.67	.79	.89	.93
10	7,10	7,10	.58	.74	.53	.60	.74	.83	.89	.96
11	7,11	7,11	.27	.47	.31	.44	.59	.70	.85	.90
12	7,12	7,12	.17	.27	.22	.27	.50	.55	.77	.81
13	7,13	7,13	.25	.41	.22	.34	.50	.67	.78	.89
14	7,14	7,14	.19	.48	.40	.58	.71	.81	.89	.95
15	7,15	7,15	.13	.26	.29	.31	.60	.64	.87	.86
16	7,16	7,16	.33	.48	.26	.34	.55	.63	.80	.85
17	7,17	7,17	.14	.25	.20	.23	.48	.50	.74	.76
18	4,15	4,15	.18	.67	.22	.45	.52	.77	.79	.93
19	4,16	4,16	.19	.23	.25	.40	.53	.66	.77	.85
20	4,17	4,17	.37	.61	.25	.40	.56	.73	.81	.92
21	4,18	4,18	.21	.32	.23	.32	.51	.66	.75	.87
22	4,19	4,19	.18	.25	.21	.27	.48	.60	.75	.81
23	4,20	4,10	.26	.52	.21	.31	.50	.70	.79	.89
24	5,10	5,10	.10	.23	.24	.29	.52	.55	.76	.78
25	5,11	5,11	.25	.28	.32	.35	.60	.61	.83	.83
26	5,12	5,12	.39	.69	.32	.45	.63	.74	.86	.93
27	5,13	5,13	.42	.68	.43	.56	.68	.78	.87	.93
28	5,14	5,14	.32	.61	.33	.42	.64	.76	.85	.93
29	8,1	8,1	.94	.96	.74	.78	.90	.91	.98	.99
30	8,2	8,2	.83	.90	.69	.77	.86	.93	.96	.99
31	8,4	8,4	.41	.58	.41	.53	.66	.79	.87	.94
32	8,5	8,5	.41	.74	.43	.62	.72	.83	.90	.96
33	8,6	8,6	.36	.55	.37	.50	.65	.79	.83	.95

Table 77. Actual p-Values and Second Set of Judges' Ratings of Items Common to the Grades 8 and 12 NAEP Test Booklets--Continued

	Placement		Actual p-Value		Judges' Item Ratings					
	Grade		Grade		Basic		Proficient		Advanced	
Common Item	8	12	8	12	8	12	8	12	8	12
34	8,7	8,7	.41	.70	.39	.55	.67	.85	.88	.97
35	8,8	8,8	.34	.49	.30	.39	.58	.71	.81	.91
36	8,9	8,9	.11	.27	.30	.37	.56	.68	.77	.89
37	8,10	8,10	.16	.30	.25	.35	.54	.66	.76	.87
38	8,11	8,11	.17	.33	.18	.26	.46	.61	.71	.85
Means:			.37	.54	.36	.47	.63	.73	.85	.91

NOTE: 38 common items.

Table 78. Summary of Achievement Levels

			Judges' Ratings		
Grade	Level	N	First	Second	Final
4	Basic	65	46.2	41.8	45.0
	Proficient	65	71.0	63.3	68.0
	Advanced	65	88.9	84.9	86.7
8	Basic	73	48.5	45.9	48.0
	Proficient	73	74.0	70.7	72.1
	Advanced	73	91.1	89.0	89.0
12	Basic	73	46.8	44.2	46.6
	Proficient	73	73.9	70.7	72.6
	Advanced	73	91.2	90.0	88.4

Table 79. Summary of Achievement Levels

		Grade 4			Grade 8			Grade 12		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
	B1	30	57.2	16.2	29	53.3	16.2	32	47.9	15.0
	B2	30	55.2	13.6	29	53.7	14.9	32	48.3	11.3
3	P1	30	78.6	13.9	29	78.1	9.4	32	77.9	11.5
	P2	30	76.9	11.5	29	77.0	9.4	32	76.1	9.2
	A1	30	92.5	8.5	29	93.2	5.3	32	94.3	4.2
	A2	30	91.0	8.0	29	92.2	5.6	32	93.2	4.4
	B1	25	35.4	11.3	32	46.8	17.3	31	53.6	13.2
	B2	25	31.4	11.5	32	44.6	14.5	31	53.5	11.8
4	P1	25	63.5	12.4	32	71.5	12.4	31	77.6	10.2
	P2	25	58.0	13.7	32	68.3	11.5	31	76.3	9.9
	A1	25	84.0	6.8	32	88.9	6.8	31	92.0	6.9
	A2	25	78.5	10.5	32	86.9	7.4	31	90.8	6.9
	B1	30	44.8	17.7	29	54.5	13.1	30	47.8	16.5
	B2	30	37.2	15.8	29	47.6	13.3	30	43.5	15.9
5	P1	30	70.0	15.1	29	77.5	8.3	30	73.7	13.3
	P2	30	61.6	15.0	29	72.4	8.4	30	69.0	14.1
	A1	30	88.9	7.7	29	92.4	4.9	30	91.5	5.5
	A2	30	82.4	9.3	29	89.5	4.5	30	87.5	8.4
	B1	25	42.2	18.3	28	53.3	14.3	29	47.9	13.9
	B2	25	33.2	16.2	28	51.8	12.8	29	44.6	13.5
6	P1	25	70.8	14.5	28	76.6	8.9	29	76.1	10.8
	P2	25	61.5	12.4	28	73.9	9.2	29	72.4	11.7
	A1	25	89.0	7.6	28	92.5	4.7	29	91.6	6.8
	A2	25	82.1	8.5	28	91.3	5.2	29	89.6	7.6
	B1	23	42.3	16.6	28	42.8	12.4	29	46.9	18.9
	B2	23	43.0	13.9	28	37.8	11.2	29	43.1	16.0
7	P1	23	66.3	15.7	28	71.5	9.0	29	72.8	13.3
	P2	23	67.2	12.6	28	65.4	10.4	29	69.2	11.0
	A1	23	86.3	9.8	28	90.9	4.8	29	90.4	7.3
	A2	23	86.5	8.7	28	86.9	5.8	29	88.5	6.7
	B1	33	50.4	19.0	30	47.3	12.5	31	46.4	17.9
	B2	33	46.8	16.8	30	44.5	10.8	31	43.4	15.1
8	P1	33	72.7	15.7	30	73.5	8.9	31	75.2	12.7
	P2	33	68.9	14.1	30	69.4	10.0	31	71.6	12.1
	A1	33	90.3	9.4	30	86.8	7.0	31	92.2	6.0
	A2	33	87.0	10.6	30	86.9	7.9	31	89.8	7.0
	B1	29	48.4	15.8	31	46.0	15.6	28	37.7	17.9
	B2	29	41.8	15.0	31	43.2	15.1	28	32.9	16.9
9	P1	29	73.1	11.3	31	73.8	12.2	28	64.4	16.1
	P2	29	65.7	13.4	31	70.4	11.3	28	59.9	16.6
	A1	29	90.6	7.2	31	91.4	6.9	28	85.5	8.5
	A2	29	84.2	11.9	31	90.0	6.2	28	81.8	10.7
	B	65	46.2	11.8	69	48.8	7.8	70	46.1	10.3
FINAL	P	65	69.6	10.2	69	72.8	5.8	70	72.1	8.3
	A	65	87.1	6.0	69	89.1	5.6	70	89.2	5.1

Table 80. Summary of Grade 4 Achievement Levels by Booklet and Round

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
	$\bar{x}$	54.0	56.3	73.4	74.2	91.1	90.0
3	SD	18.6	17.7	16.3	11.9	11.9	12.6
	$\bar{x}$	46.1	48.9	67.3	69.4	86.4	86.9
7	SD	15.0	14.7	17.4	14.6	11.6	11.4
	$\bar{x}$	50.2	50.8	69.2	69.7	87.9	87.0
8	SD	15.6	16.7	17.0	16.0	12.7	13.2
	$\bar{x}$	51.6		72.2		87.8	
Final	SD	13.2		15.2		11.4	

NOTE: Booklet = 15c Blocks = 3,7,8 Judges = 10

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
	$\bar{x}$	48.9	48.9	76.3	76.5	89.9	90.1
3	SD	15.3	12.3	14.8	9.1	8.9	5.1
	$\bar{x}$	36.4	32.8	66.6	63.0	83.8	80.3
4	SD	11.9	12.4	14.0	11.2	7.9	5.5
	$\bar{x}$	42.8	34.6	72.9	64.8	87.9	80.6
6	SD	13.5	16.4	11.6	8.9	6.1	3.5
	$\bar{x}$	42.5		68.9		85.9	
Final	SD	6.9		5.1		3.5	

NOTE: Booklet = 11r Blocks = 3,4,6 Judges = 8

Table 80. Summary of Grade 4 Achievement Levels by Booklet and Round -- Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
6	$\bar{x}$	42.5	32.8	70.0	63.0	91.0	86.0
	SD	19.9	10.7	12.3	8.0	6.3	3.8
7	$\bar{x}$	46.0	40.7	73.0	71.7	90.3	89.3
	SD	19.6	10.0	15.0	6.1	7.3	4.6
9	$\bar{x}$	45.7	38.7	72.8	66.5	91.2	87.3
	SD	21.4	13.9	12.6	8.0	4.3	5.2
Final	$\bar{x}$	40.8		69.2		87.5	
	SD	8.1		4.4		2.8	

NOTE: Booklet = 14cr Blocks = 6,7,9 Judges = 6

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	36.5	31.9	64.5	56.0	85.7	77.5
	SD	9.4	12.1	12.1	17.0	7.3	15.2
8	$\bar{x}$	49.5	46.6	72.9	68.3	90.7	86.8
	SD	18.3	15.6	12.1	14.0	6.4	11.4
9	$\bar{x}$	46.2	40.1	69.5	62.0	87.5	80.2
	SD	17.1	18.3	13.2	18.2	10.1	16.7
Final	$\bar{x}$	45.1		66.2		86.6	
	SD	12.4		14.2		7.6	

NOTE: Booklet = 16C Blocks = 4,8,9 Judges = 6

Table 80. Summary of Grade 4 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	32.0	28.7	57.7	54.8	81.0	77.8
	SD	14.8	10.7	10.7	9.2	3.6	4.2
5	$\bar{x}$	36.3	30.3	62.0	53.3	85.8	76.8
	SD	14.4	11.5	13.0	12.3	5.8	7.6
7	$\bar{x}$	33.3	36.8	59.5	60.8	83.7	84.8
	SD	16.4	15.1	14.0	13.1	9.1	6.9
Final	$\bar{x}$	41.8		67.7		83.0	
	SD	4.3		6.7		4.1	

NOTE: Booklet = 12 Blocks = 4,5,7 Judges = 6

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
5	$\bar{x}$	43.5	36.8	70.5	62.9	89.5	84.1
	SD	22.4	20.2	19.1	18.5	9.7	10.2
6	$\bar{x}$	41.3	32.7	68.8	58.1	87.9	80.7
	SD	20.9	18.7	17.7	15.4	9.6	11.1
8	$\bar{x}$	51.3	43.5	75.5	68.8	91.9	87.1
	SD	23.3	18.6	18.1	13.7	8.8	8.1
Final	$\bar{x}$	42.2		67.4		87.2	
	SD	14.0		10.5		2.4	

NOTE: Booklet = 13CR Blocks = 5,6,8 Judges = 12

Table 80. Summary of Grade 4 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
	$\overline{x}$	65.3	58.5	84.4	79.5	95.5	92.5
3	SD	11.4	9.5	8.3	7.8	3.2	4.1
	$\overline{x}$	50.3	41.2	73.6	64.3	89.8	83.4
5	SD	12.7	12.1	10.6	11.8	6.5	8.7
	$\overline{x}$	51.9	43.0	76.4	68.6	93.2	86.4
9	SD	12.0	13.2	8.4	10.4	3.9	8.2
	$\overline{x}$	54.1		74.4		89.4	
Final	SD	10.7		5.0		2.2	

NOTE: Booklet = 17C Blocks = 3,5,9 Judges = 12

Table 81. Summary of Grade 8 Achievement Levels by Booklet and Round

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
5	$\bar{x}$	53.0	49.0	75.3	71.8	91.1	89.3
	SD	13.5	13.7	8.7	9.1	5.5	5.5
6	$\bar{x}$	57.7	55.1	78.7	75.6	92.9	90.9
	SD	13.7	12.5	7.8	7.7	4.8	4.1
8	$\bar{x}$	48.7	44.4	71.4	66.0	87.9	83.7
	SD	14.0	11.7	10.6	10.3	8.6	8.3
Final	$\bar{x}$	51.3		73.1		89.2	
	SD	6.9		5.7		4.9	

NOTE: Booklet = 10cp Blocks = 5,6,8 Judges = 9

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
3	$\bar{x}$	42.5	44.3	70.0	68.7	89.1	88.0
	SD	13.5	11.4	7.5	7.1	6.6	6.6
4	$\bar{x}$	38.0	36.8	64.3	61.0	86.3	83.7
	SD	12.4	11.4	9.0	10.3	6.9	7.7
6	$\bar{x}$	48.4	46.7	73.0	70.3	91.2	89.4
	SD	14.7	13.9	8.0	9.9	5.5	6.3
Final	$\bar{x}$	44.7		68.5		88.7	
	SD	8.7		5.7		5.0	

NOTE: Booklet = 8P Blocks = 3,4,6 Judges = 10

Table 81. Summary of Grade 8 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	43.6	42.9	67.8	67.3	87.7	87.6
	SD	18.5	12.5	12.5	6.4	6.8	4.4
5	$\bar{x}$	54.5	46.9	78.1	72.4	92.7	89.8
	SD	15.0	11.1	9.6	7.8	6.0	4.5
7	$\bar{x}$	45.0	38.8	72.1	66.2	91.5	86.9
	SD	12.2	12.3	9.8	11.1	4.7	5.5
Final	$\bar{x}$	45.9		70.9		89.5	
	SD	4.7		3.1		2.6	

NOTE: Booklet = 9 Blocks = 4,5,7 Judges = 10

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
3	$\bar{x}$	55.7	57.9	81.0	80.8	94.9	95.1
	SD	18.8	12.9	7.0	6.5	4.6	3.6
7	$\bar{x}$	40.0	35.9	72.8	65.6	91.9	86.7
	SD	15.6	13.5	4.9	8.5	4.5	7.7
8	$\bar{x}$	44.9	42.1	74.6	70.3	92.3	88.8
	SD	14.9	11.5	7.1	7.9	5.9	7.7
Final	$\bar{x}$	49.3		74.8		91.3	
	SD	4.9		3.3		2.2	

NOTE: Booklet = 12C Blocks = 3,7,8 Judges = 9

Table 81. Summary of Grade 8 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
6	$\bar{x}$	54.4	54.1	78.4	76.2	93.6	93.7
	SD	14.3	11.5	9.9	9.5	3.7	4.4
7	$\bar{x}$	43.2	38.7	69.4	64.3	89.3	87.1
	SD	9.5	7.8	11.7	12.3	5.3	4.3
9	$\bar{x}$	50.4	46.2	74.9	70.0	93.6	91.8
	SD	14.2	10.8	13.1	10.4	3.3	4.5
Final	$\bar{x}$	50.7		74.7		91.0	
	SD	6.4		4.2		2.3	

NOTE: Booklet = 11CP Blocks = 6,7,9 Judges = 9

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	56.7	52.4	80.5	75.1	92.0	89.0
	SD	16.0	15.2	9.6	12.4	5.8	6.7
8	$\bar{x}$	48.1	46.3	74.3	71.3	89.9	87.6
	SD	10.1	10.1	9.3	11.2	6.4	7.6
9	$\bar{x}$	46.3	44.3	75.6	72.8	91.7	90.3
	SD	13.6	14.	9.4	10.7	6.0	6.8
Final	$\bar{x}$	49.9		74.2		89.9	
	SD	8.6		8.6		4.9	

NOTE: Booklet = 13C Blocks = 4,8,9 Judges = 12

Table 81. Summary of Grade 8 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
3	$\overline{x}$	61.9	59.2	83.5	81.8	95.7	93.8
	SD	14.6	16.2	7.7	8.5	1.9	3.4
5	$\overline{x}$	55.8	46.9	78.9	73.1	93.2	89.3
	SD	12.0	16.	6.9	9.1	2.9	4.0
9	$\overline{x}$	41.6	39.1	70.7	67.9	89.1	88.1
	SD	19.1	19.5	14.7	13.3	9.6	7.1
Final	$\overline{x}$	49.9		73.7		84.6	
	SD	11.5		5.4		10.6	

NOTE: Booklet = 14C Block = 3,5,9 Judges = 10

Table 82. Summary of Grade 12 Achievement Levels by Booklet and Round

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	52.9	53.4	75.0	74.0	91.2	90.1
	SD	12.7	11.1	7.6	4.4	4.6	3.5
5	$\bar{x}$	50.1	45.4	74.5	70.0	92.4	89.7
	SD	15.6	12.9	7.8	7.6	4.0	3.9
7	$\bar{x}$	47.7	43.7	75.0	70.3	92.4	89.6
	SD	19.6	16.4	10.3	9.9	3.3	4.3
Final	$\bar{x}$	50.2		72.6		88.9	
	SD	4.8		5.5		2.9	

NOTE: Booklet = 9 Blocks = 4,5,7 Judges = 10

Table 82. Summary of Grade Achievement Levels By Booklet and Round -- Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
3	$\bar{x}$	42.9	46.0	74.1	75.0	93.5	93.4
	SD	8.9	9.4	5.5	6.5	4.3	4.7
4	$\bar{x}$	46.8	46.5	75.5	73.9	91.4	90.5
	SD	6.7	5.4	6.4	7.2	5.5	6.1
6	$\bar{x}$	39.4	36.9	71.4	66.7	89.9	87.5
	SD	7.5	6.9	9.6	10.8	8.4	8.7
Final	$\bar{x}$	44.4		73.3		90.7	
	SD	3.9		4.8		4.7	

NOTE: Booklet = 8 Blocks = 3,4,6 Judges = 11

Table 82. Summary of Grade 12 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
	$\bar{x}$	54.0	47.9	76.1	71.4	91.3	87.8
6	SD	11.7	12.4	11.8	11.7	7.0	7.9
	$\bar{x}$	56.6	49.6	76.8	70.6	91.1	87.9
7	SD	19.2	18.5	13.8	14.5	7.1	7.8
	$\bar{x}$	45.8	39.5	67.3	61.9	84.5	81.3
9	SD	18.7	19.9	15.5	16.4	9.2	9.7
	$\bar{x}$	44.0		68.3		85.9	
Final	SD	18.6		13.1		7.6	

NOTE: Booklet = 11C Blocks = 6,7,9 Judges = 8)

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
5	$\bar{x}$	55.6	54.1	80.8	79.4	93.2	92.4
	SD	12.6	13.5	9.2	9.6	4.9	6.3
	$\bar{x}$	52.4	50.5	81.4	79.6	93.8	93.2
6	SD	17.0	16.5	9.5	9.5	4.3	5.0
	$\bar{x}$	55.1	51.3	82.1	78.5	94.2	92.3
8	SD	17.6	16.6	8.3	9.4	4.2	5.7
	$\bar{x}$	53.3		78.2		92.5	
Final	SD	11.3		8.1		4.2	

NOTE: Booklet = 10C Blocks = 5,6,8 Judges = 10)

Table 82. Summary of Grade 12 Achievement Levels by Booklet and Round -- Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
4	$\bar{x}$	61.7	61.1	82.6	81.3	93.5	91.9
	SD	15.4	13.5	14.1	14.5	9.9	10.1
8	$\bar{x}$	43.3	41.9	71.5	70.0	91.6	89.9
	SD	19.5	15.9	13.0	14.5	6.3	8.5
9	$\bar{x}$	40.1	35.6	70.9	66.9	89.9	87.0
	SD	18.2	16.4	14.6	16.3	7.8	11.4
Final	$\bar{x}$	45.5		73.0		88.4	
	SD	9.8		6.4		5.2	

NOTE: Booklet = 13C Blocks = 4,8,9 Judges = 10)

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
3	$\bar{x}$	53.6	53.1	83.7	81.0	95.3	94.5
	SD	14.5	9.6	8.7	6.9	4.5	4.3
7	$\bar{x}$	39.3	37.9	68.0	67.2	88.2	88.0
	SD	16.0	13.0	14.9	9.7	9.8	8.0
8	$\bar{x}$	41.3	37.5	72.3	66.9	90.8	87.4
	SD	15.0	10.2	14.0	9.7	6.9	6.2
Final	$\bar{x}$	46.3		74.9		92.1	
	SD	7.3		1.3		1.9	

NOTE: Booklet = 12C Blocks = 3,7,8 Judges = 11

Table 82. Summary of Grade 12 Achievement Levels by Booklet and Round--Continued

		Basic		Proficient		Advanced	
Block		1	2	1	2	1	2
	$\overline{x}$	47.1	45.5	75.6	72.0	94.2	91.7
3	SD	19.5	14.0	16.5	11.8	3.9	4.1
	$\overline{x}$	37.8	30.9	65.8	57.7	88.8	80.3
5	SD	17.0	12.9	17.4	15.1	6.7	8.9
	$\overline{x}$	28.8	24.9	55.5	51.2	82.0	77.0
9	SD	14.2	12.7	15.4	14.6	7.4	9.4
	$\overline{x}$	38.4		63.4		84.5	
Final	SD	8.6		8.8		4.0	

NOTE: Booklet = 14C Blocks = 3,5,9 Judges = 10

Table 83. Grade 4 Achievement Levels by State

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
3	Basic1	7	52.1	13.4	10	57.9	14.3	7	62.4	16.9	6	55.6	22.7
	Basic2	7	50.7	9.9	10	57.8	9.8	7	59.7	14.5	6	50.8	20.8
	Proficient1	7	73.7	11.7	10	80.7	10.3	7	79.8	13.5	6	79.1	22.5
	Proficient2	7	74.0	6.8	10	79.1	6.3	7	76.7	10.9	6	77.0	21.8
	Advanced1	7	91.0	6.9	10	95.0	2.8	7	91.8	8.4	6	91.0	15.8
	Advanced2	7	92.2	4.7	10	92.5	3.7	7	89.2	6.3	6	89.1	15.9
4	Basic1	9	30.7	10.6	5	37.6	5.3	7	42.5	12.4	4	30.2	12.3
	Basic2	9	27.6	9.9	5	33.6	6.0	7	40.1	13.2	4	21.7	7.1
	Proficient1	9	56.8	8.9	5	70.2	7.5	7	66.0	13.4	4	65.7	19.2
	Proficient2	9	50.4	10.5	5	66.8	3.8	7	62.7	15.4	4	55.5	19.0
	Advanced1	9	80.1	4.0	5	86.8	4.3	7	85.0	8.4	4	87.5	9.3
	Advanced2	9	72.4	12.8	5	84.2	3.7	7	81.7	10.1	4	79.2	5.5
5	Basic1	8	30.3	10.8	8	45.1	8.6	8	48.3	18.5	6	58.6	22.2
	Basic2	8	25.2	9.8	8	34.0	5.1	8	43.6	17.5	6	49.0	19.4
	Proficient1	8	59.3	11.4	8	73.0	8.6	8	70.3	14.1	6	79.8	21.3
	Proficient2	8	50.7	12.2	8	61.5	9.9	8	64.5	14.1	6	72.1	18.9
	Advanced1	8	88.0	4.2	8	88.8	6.7	8	88.2	6.2	6	91.0	14.1
	Advanced2	8	81.2	8.9	8	79.2	6.5	8	83.2	8.4	6	86.8	13.6
6	Basic1	7	36.5	13.3	5	50.4	16.7	5	48.2	16.3	8	38.3	23.8
	Basic2	7	27.8	13.7	5	39.8	5.1	5	44.4	16.7	8	26.7	19.2
	Proficient1	7	64.7	12.8	5	76.4	13.4	5	69.6	7.9	8	73.3	19.3
	Proficient2	7	56.4	11.7	5	68.0	11.5	5	64.4	10.0	8	60.0	14.5
	Advanced1	7	87.7	4.7	5	90.4	6.0	5	86.8	4.4	8	90.7	11.8
	Advanced2	7	81.0	8.3	5	83.6	12.2	5	83.8	6.2	8	81.1	8.7
7	Basic1	8	40.8	16.2	7	42.4	16.0	4	40.7	13.5	4	46.7	26.0
	Basic2	8	39.8	14.6	7	52.5	11.1	4	42.7	10.8	4	33.0	13.4
	Proficient1	8	67.5	15.0	7	65.0	16.4	4	62.2	9.2	4	70.5	24.8
	Proficient2	8	65.2	14.4	7	73.7	7.4	4	63.0	5.5	4	63.7	19.9
	Advanced1	8	88.1	8.3	7	86.2	7.9	4	81.5	7.0	4	87.7	17.9
	Advanced2	8	86.5	8.8	7	89.7	4.5	4	83.0	3.2	4	84.5	16.8
8	Basic1	8	45.0	14.9	9	54.5	18.3	8	54.8	18.2	8	46.5	24.7
	Basic2	8	42.0	13.8	9	53.0	13.9	8	51.3	17.4	8	39.8	20.6
	Proficient1	8	71.1	13.1	9	76.5	14.0	8	75.1	13.4	8	67.6	22.1
	Proficient2	8	64.7	11.9	9	76.0	9.5	8	73.2	13.3	8	60.6	17.3
	Advanced1	8	90.0	6.4	9	92.8	5.2	8	90.8	6.8	8	87.1	16.3
	Advanced2	8	83.7	12.9	9	90.7	7.4	8	89.8	7.0	8	83.0	13.3

Table 83. Grade 4 Achievement Levels by State--Continued

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
	Basic1	7	42.2	6.3	10	43.1	7.0	6	59.6	22.2	6	53.3	22.1
	Basic2	7	34.0	9.5	10	37.8	8.2	6	56.0	20.3	6	39.5	16.2
	Proficient1	7	66.5	7.0	10	72.3	7.4	6	79.6	15.9	6	75.3	13.8
9	Proficient2	7	57.0	14.5	10	66.9	7.1	6	75.8	15.7	6	63.5	13.6
	Advanced1	7	85.1	9.2	10	90.7	3.4	6	93.6	9.4	6	93.8	3.8
	Advanced2	7	74.0	17.2	10	86.4	6.8	6	90.0	9.9	6	86.8	7.5
	Basic	18	38.0	9.5	18	49.7	6.7	15	54.0	11.0	14	44.0	13.9
Final	Proficient	18	64.0	8.9	18	74.0	5.9	15	72.6	9.1	14	67.7	13.9
	Advanced	18	85.5	5.1	18	88.3	3.6	15	88.0	5.3	14	86.3	9.3

Table 84. Grade 8 Achievement Levels by State

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
3	Basic1	7	59.0	8.3	7	62.4	19.2	7	50.4	14.7	8	42.8	19.5
	Basic2	7	58.4	8.4	7	61.6	19.2	7	50.0	14.8	8	45.8	12.2
	Proficient1	7	79.6	5.4	7	82.1	13.8	7	74.1	10.0	8	76.6	6.3
	Proficient2	7	78.7	5.4	7	80.6	14.9	7	75.0	9.2	8	74.0	6.1
	Advanced1	7	92.6	4.7	7	93.0	7.7	7	91.6	6.2	8	95.3	1.3
	Advanced2	7	92.3	3.5	7	91.6	9.4	7	91.6	6.2	8	93.3	2.3
4	Basic1	7	48.3	14.5	10	50.8	22.0	7	41.7	14.9	8	44.8	16.8
	Basic2	7	47.4	12.9	10	48.3	17.5	7	39.6	10.3	8	41.8	14.9
	Proficient1	7	70.7	9.4	10	72.7	15.3	7	68.6	8.3	8	73.1	15.2
	Proficient2	7	69.7	8.6	10	70.3	11.8	7	63.4	7.5	8	68.6	16.3
	Advanced1	7	86.0	5.3	10	88.3	7.2	7	85.1	6.1	8	95.3	2.8
	Advanced2	7	85.9	4.8	10	86.5	9.3	7	82.1	6.4	8	92.5	4.0
5	Basic1	7	47.7	14.4	7	53.3	16.2	7	55.6	11.5	8	60.5	9.2
	Basic2	7	44.0	14.4	7	50.7	18.5	7	49.1	12.4	8	46.5	8.5
	Proficient1	7	70.4	8.4	7	78.7	8.7	7	79.6	5.1	8	80.9	7.6
	Proficient2	7	68.6	10.0	7	75.7	10.7	7	75.7	5.1	8	70.1	5.7
	Advanced1	7	89.0	4.7	7	93.0	5.2	7	94.3	2.8	8	93.1	5.5
	Advanced2	7	88.0	5.4	7	91.7	5.7	7	90.9	2.3	8	87.6	3.4
6	Basic1	6	64.2	14.8	9	50.0	15.2	6	49.3	12.0	7	51.7	12.4
	Basic2	6	60.2	12.5	9	50.0	13.5	6	51.5	10.4	7	47.1	13.3
	Proficient1	6	85.2	6.2	9	74.8	9.0	6	71.5	8.7	7	75.9	6.9
	Proficient2	6	81.8	4.5	9	72.9	9.9	6	73.0	8.1	7	69.1	9.3
	Advanced1	6	95.8	2.4	9	90.3	5.3	6	89.0	4.3	7	95.1	2.4
	Advanced2	6	95.0	2.2	9	89.0	5.9	6	89.5	6.1	7	92.4	4.0
7	Basic1	8	45.8	12.1	7	38.3	10.6	6	48.2	12.7	7	39.4	13.8
	Basic2	8	40.0	11.6	7	37.0	9.8	6	44.8	9.2	7	30.1	10.5
	Proficient1	8	67.6	8.9	7	69.4	11.1	6	74.8	9.7	7	75.0	4.8
	Proficient2	8	62.0	7.4	7	69.0	10.1	6	72.0	9.2	7	60.0	11.8
	Advanced1	8	88.0	4.2	7	89.6	6.4	6	93.7	2.3	7	93.3	3.2
	Advanced2	8	84.5	3.4	7	88.0	4.1	6	91.7	3.0	7	84.1	8.5

Table 84. Grade 8 Achievement Levels by State--Continued

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
8	Basic1	7	50.1	12.8	10	47.7	9.8	7	46.1	7.7	6	44.7	21.0
	Basic2	7	46.6	10.0	10	43.2	11.1	7	44.3	6.9	6	44.3	16.3
	Proficient1	7	73.6	6.7	10	74.6	8.7	7	69.1	5.8	6	76.8	13.8
	Proficient2	7	70.9	6.5	10	69.9	10.6	7	65.7	7.0	6	71.2	15.4
	Advanced1	7	89.4	6.2	10	91.4	6.8	7	87.4	4.7	6	91.5	10.4
	Advanced2	7	88.6	6.3	10	88.3	7.0	7	84.1	6.3	6	85.2	12.4
9	Basic1	6	52.8	16.5	10	49.8	18.0	8	34.0	12.6	7	48.3	7.3
	Basic2	6	47.7	15.6	10	50.0	18.2	8	32.3	10.0	7	42.1	8.5
	Proficient1	6	77.2	9.7	10	77.4	11.5	8	63.6	14.0	7	77.4	6.9
	Proficient2	6	70.3	10.3	10	76.9	9.2	8	62.1	11.5	7	70.7	10.5
	Advanced1	6	92.0	4.7	10	92.7	4.0	8	86.6	10.8	7	94.4	3.6
	Advanced2	6	88.7	5.6	10	92.7	5.4	8	86.0	6.8	7	91.9	5.8
FINAL	Basic	16	51.7	3.8	20	51.7	9.7	16	45.7	7.3	17	45.5	6.6
	Proficient	16	73.4	3.2	20	75.2	6.3	16	70.8	6.2	17	71.4	6.0
	Advanced	16	89.1	2.9	20	88.5	8.9	16	87.9	4.0	17	91.2	3.0

Table 85. Grade 12 Achievement Levels by State

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
3	Basic1	9	47.9	13.5	6	55.8	18.3	9	41.0	15.6	8	50.4	12.6
	Basic2	9	47.4	13.4	6	55.3	11.7	9	46.9	11.0	8	45.5	8.3
	Proficient1	9	78.0	8.8	6	80.8	11.6	9	72.6	14.9	8	81.5	9.2
	Proficient2	9	75.2	11.1	6	79.5	8.2	9	73.4	8.8	8	77.6	8.6
	Advanced1	9	93.0	4.2	6	95.2	4.9	9	94.7	4.4	8	94.8	3.8
	Advanced2	9	91.2	5.2	6	94.7	3.6	9	94.0	4.5	8	93.5	3.8
4	Basic1	9	53.7	12.5	8	58.1	18.0	9	51.2	5.9	5	50.4	17.0
	Basic2	9	53.1	9.9	8	57.3	16.2	9	53.4	6.4	5	48.0	15.2
	Proficient1	9	76.9	11.4	8	81.1	12.7	9	75.7	8.3	5	77.0	7.2
	Proficient2	9	74.8	9.3	8	78.6	14.2	9	76.0	8.9	5	76.0	6.4
	Advanced1	9	89.4	10.0	8	93.6	5.7	9	92.3	5.5	5	93.4	3.8
	Advanced2	9	88.0	8.2	8	92.4	6.9	9	90.9	6.7	5	93.4	3.8
5	Basic1	9	49.0	14.9	5	50.0	10.7	9	53.7	18.7	7	37.3	17.1
	Basic2	9	42.2	13.3	5	47.6	9.2	9	49.0	19.3	7	35.0	17.1
	Proficient1	9	75.3	9.7	5	73.2	6.1	9	74.6	20.8	7	70.9	10.8
	Proficient2	9	68.6	13.5	5	71.8	7.9	9	69.0	19.9	7	67.7	11.7
	Advanced1	9	91.0	4.8	5	92.6	2.8	9	92.3	6.1	7	90.1	7.4
	Advanced2	9	85.2	8.8	5	91.2	2.8	9	87.3	9.9	7	87.9	8.9
6	Basic1	8	42.1	9.6	7	43.1	12.5	8	57.9	13.7	6	47.8	16.4
	Basic2	8	38.6	7.4	7	42.3	10.8	8	56.6	14.1	6	39.3	14.1
	Proficient1	8	68.8	9.0	7	73.0	10.4	8	85.5	7.5	6	77.2	9.2
	Proficient2	8	63.6	7.4	7	70.4	9.1	8	84.3	9.3	6	70.8	10.5
	Advanced1	8	85.4	7.5	7	91.3	5.8	8	96.9	2.6	6	93.3	4.6
	Advanced2	8	82.0	6.1	7	90.0	6.0	8	96.4	3.1	6	90.0	7.0
7	Basic1	8	47.1	10.6	6	43.0	14.9	8	51.5	25.8	7	44.9	22.8
	Basic2	8	43.8	6.9	6	41.8	12.8	8	51.3	22.6	7	34.3	15.0
	Proficient1	8	71.5	13.1	6	69.7	14.5	8	72.6	17.3	7	77.3	7.9
	Proficient2	8	68.5	10.9	6	67.5	11.6	8	72.4	14.1	7	68.0	7.7
	Advanced1	8	88.3	8.4	6	90.0	6.4	8	90.1	9.2	7	93.7	4.2
	Advanced2	8	86.8	8.2	6	88.3	4.5	8	89.5	7.9	7	89.6	5.8
8	Basic1	9	45.6	11.9	7	54.3	19.3	9	41.1	22.6	6	46.3	16.8
	Basic2	9	41.2	7.2	7	52.3	16.6	9	41.4	20.2	6	38.0	12.6
	Proficient1	9	72.1	8.5	7	80.3	12.6	9	69.4	16.3	6	82.5	7.4
	Proficient2	9	67.9	6.8	7	78.6	12.5	9	68.7	16.1	6	73.7	9.2
	Advanced1	9	88.0	5.3	7	94.1	5.9	9	92.0	5.7	6	96.3	3.9
	Advanced2	9	84.6	6.5	7	93.4	5.7	9	89.8	7.5	6	93.3	3.4

Table 85. Grade 12 Achievement Levels by State--Continued

		CT			MI			CA			FL		
Block	Level	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD	N	$\bar{x}$	SD
	Basic1	8	30.8	12.2	6	49.0	21.2	8	37.1	21.2	6	36.3	14.3
	Basic2	8	30.0	11.4	6	42.2	19.5	8	34.4	22.9	6	25.5	7.8
	Proficient1	8	57.6	13.9	6	74.5	18.2	8	62.4	18.6	6	65.8	10.6
9	Proficient2	8	55.9	13.7	6	70.7	18.3	8	57.6	21.3	6	57.3	9.0
	Advanced1	8	79.5	8.5	6	90.0	8.6	8	87.8	7.9	6	86.2	6.2
	Advanced2	8	77.3	8.2	6	89.0	9.3	8	82.1	14.4	6	80.2	7.5
	Basic	20	46.1	5.2	15	48.4	8.8	20	50.2	13.2	15	38.3	8.9
FINAL	Proficient	20	71.1	6.1	15	73.9	7.2	20	73.2	11.9	15	70.3	5.9
	Advanced	20	87.0	4.7	15	89.9	3.3	20	90.9	5.9	15	89.0	5.4

Appendix J

Setting Appropriate Achievement Levels
for the
National Assessment of Educational Progress
Policy Framework and Technical Procedures

374

TABLE OF CONTENTS

Page

Executive Summary and Board Action 333

PART 1 - POLICY FRAMEWORK

Background and Rationale 337
The Changing Environment 339
The Need for Appropriate Achievement Levels 341
Framework and Definitions 342
Procedures for Establishing Specific Achievement Level 344
Reporting NAEP in Terms of Achievement Levels 346
When Should Achievement Levels Be Set? 347
NAEP and International Achievement Levels 349
Rejected Alternative Proposals to Use NAEP for Setting
Achievement Goals 350
Endnote: The Promise and Some Cautions 353

PART 2 - TECHNICAL PROCEDURES

Introduction 354
A Modified Angoff Procedure 355
Assessment Content 356
Achievement Levels 357
Number of Levels and Sales for Each Grade 358
Procedures for Setting Achievement Levels 359
Appendices 364

PART 3 - DISPLAYING NAEP RESULTS IN TERMS OF ACHIEVEMENT LEVELS 370

Executive Summary and Board Action

Approved Unanimously May 11, 1990
At Meeting in Washington, D.C.

Setting appropriate achievement levels on the National Assessment of Educational Progress will help define some of the important outcomes of education, stating clearly what students should know and be able to do at key grades in school. This will make the Assessment far more useful to parents and policymakers as a measure of performance in American schools and perhaps as an inducement to higher achievement. The achievement levels will be used for reporting NAEP results in a way which greatly increases their value to the American public.

The National Assessment Governing Board notes its statutory responsibility to (1) take "appropriate actions...to improve the form and use of the National Assessment" and (2) identify "appropriate achievement goals for each...grade (and) subject area to be tested under the National Assessment." To carry out these responsibilities the Board shall establish appropriate achievement levels on the National Assessment and endorses in concept the accompanying Committee paper titled, Setting Appropriate Achievement Levels for the National Assessment of Educational Progress, dated May 10, 1990. Further, the Board approves the following policy framework, definitions, and technical procedures for establishing achievement levels on the National Assessment:

1. Three achievement levels with clear distinctions between them shall be established for each grade and subject tested under NAEP. These levels shall be called:

(a) **Proficient**. This central level represents solid academic performance for each grade tested--4, 8, and 12. It will reflect a consensus that students reaching this level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. At grade 12 the proficient level will encompass a body of subject-matter knowledge and analytical skills, of cultural literacy and insight, that all high school graduates should have for democratic citizenship, responsible adulthood, and productive work.

(b) **Advanced**. This higher level signifies superior performance beyond proficient grade-level mastery at grades 4, 8, and 12. For 12th grade the advanced level will show readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, it may be based in part on international comparisons of academic achievement and may also be related to Advanced Placement and other college placement exams.

(c) **Basic**. This level, below proficient, denotes partial mastery of knowledge and skills that are fundamental for proficient work at each grade--4, 8, and 12. For 12th grade this will be higher than minimum competency skills (which normally are taught in elementary and junior high schools) and will cover significant elements of standard high school-level work.

2. It is the Board's intention to use this framework of basic, proficient, and advanced achievement levels as the primary means of reporting results for all newly-developed assessments in 1992 and thereafter. The framework shall first be applied in reporting the 1990 National Assessment of mathematics, contingent upon the successful conduct of the process to set achievement levels adopted by the Board. If the process is carried out successfully, results in terms of three achievement levels per grade shall be a prominent part of the initial release of national data from the 1990 math assessment. In the simultaneous release of data from the trial state assessment of 8th grade math, each state will have the option of having its results displayed in terms of the three achievement levels in addition to the previously-developed formats of five across-grade distributional proficiency levels, quartiles, and percent of correct answers. With the assistance of the states, the several ways of reporting results from the trial state assessment shall be evaluated.

3. The process for determining achievement levels shall be a logical continuation of the national consensus effort used in developing the content and objectives of the National Assessment.

4. To assist in defining achievement levels for the 1990 assessment of mathematics the Board shall appoint an ad hoc advisory panel, divided into separate subcommittees for grades 4, 8 and 12. The panel will be broadly representative and will consist of state and local educators, scholars, employers, civic group representatives, and other interested citizens.

5. The subcommittees will be charged with using a proven judgment procedure to recommend which test questions and/or which proportion of questions students need to answer correctly to reach various achievement levels in accordance with this framework. As part of its deliberations, the panel will be required to prepare detailed descriptions of the subject-matter knowledge and skills proposed for each achievement level. These shall be illustrated by representative sample items and scoring protocols.

6. In preparing descriptions of achievement levels and assigning test items to them the panel members shall use their best judgment and expertise and shall also take into account a wide range of background information and frames of reference. These may include relevant curriculum and testing data from state, local, national, and international levels; comments solicited from interested citizens, specialists, and education agencies; research on the performance of different groups, such as college students and other young adults; or studies equating NAEP with other testing programs. Specifically, the panel may consider data from the 1988 International Assessment of Mathematics and Science and from Advanced Placement examinations. The panel shall refer to sources such as these in presenting the rationale for the proposed achievement levels. The panel shall ensure coherence and consistency in the recommended achievement levels over the three grades.

7. The panel shall submit proposed descriptions of mathematics achievement levels to the Board by September 20, 1990. Its report shall include sample questions, justification for the levels proposed, and a full explanation of its procedures.

8. The Board shall seek public comment on the panel's recommendations and shall hold a public forum on them during October 1990. The Board's schedule calls for it to take action on the mathematics achievement levels during its meeting of November 16 and 17.

9. It is the Board's intention that both state and national data for the 1992 assessments shall be reported initially and primarily in terms of achievement levels and that this shall be made known to the states as an element of the 1992 trial state assessment. The Board's process for establishing achievement levels will be revised as necessary on the basis of experience and practicality.

10. The Board shall ensure that all newly-developed NAEP assessments contain a broad range of content so that three achievement levels can be established for each grade in accordance with Board policy. In addition, the consensus process for developing objectives and specifications for any future assessment shall consider the three achievement levels per grade and the possibility of grade-specific scales.

11. The 1990 assessments shall continue the practice of reporting NAEP data for each subject on a common across-grade scale that spans grades, 4, 8, and 12. However, the Board is concerned that such scaling may not adequately show variations of performance within each grade. The Board intends to continue to explore the issue of grade-specific and across-grade scales. It intends to reach a decision on which scale or scales shall be used for reporting the 1992 and subsequent assessments. A timeline for making this decision shall be developed by NAGB staff, in consultation with NCES and ETS, for consideration by the Board at its August 1990 meeting.

Part 1

Policy Framework

Background and Rationale

Among the most significant responsibilities of the National Assessment Governing Board are (1) "taking appropriate actions... to improve the form and use of the National Assessment" and (2) setting "appropriate achievement goals" for each grade and subject tested under NAEP. The two responsibilities fit well together. By defining levels of appropriate achievement on the National Assessment the Board will increase greatly the significance and usefulness of NAEP results to educators, policymakers, and the American public.

The statute (P.L. 100-297) creating the Board assigns to it certain explicit responsibilities:

- "Taking appropriate actions needed to improve the form and use of the National Assessment;
- "Developing...standards for analysis plans and for reporting and disseminating (NAEP) results;
- "Developing standards and procedures for interstate, regional, and national comparisons;
- "Identifying appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;
- "Developing assessment objectives (and) specifications;"
- Devising goal statements for each learning area assessment "through a national consensus approach that provides for the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the general public."

The National Assessment Governing Board is not authorized to establish any overarching national goals for education. It does have authority to define levels of achievement that will serve as "appropriate achievement goals" on National Assessment exams. With such achievement levels defined, NAEP results will be reported in terms that better denote the quality or value of student achievement than do the numerical scores that represent the range of student performance.

By law, the National Assessment is a survey--not a mass individual testing program--in which representative samples of students are asked questions in different academic subjects. The assessment provides information on aggregate or group performance; it is forbidden by law to report data on individuals.

Hence, the achievement levels defined by the Board will be used for reporting group data and making it more meaningful. The assessment will not become a device for certifying or classifying individual students.

In a letter to the Governing Board, Education Secretary Lauro F. Cavazos said that by "setting achievement standards for the National Assessment" the Board "would fulfill (its) statutory responsibility...(under) the Hawkins-Stafford Amendments of 1988...The result would be a clear definition of what constitutes grade level performance in each subject so that future National Assessment of Educational Progress (NAEP) reports could provide data on the proportion of students who achieve that standard and in what ways American students exceed or fall short."

The Secretary concluded that such Board action "is not only in keeping with the charge of the law, but is a constructive and complementary addition...to the work of the President and the Governors as they establish goals for performance of the Nation's education system." (Cavazos letter of Jan. 24, 1990)

The Changing Environment

When the U.S. Office of Education was created in 1867, Congress charged it with the duty of "collecting such statistics and facts as shall show the condition and progress of education in the several states." Over the ensuing century the Office collected a great deal of information about school attendance, spending, class size, and graduates; it reported virtually nothing about what students had learned.

It was not until the mid-1960s that President Johnson and U.S. Commissioner of Education Francis Keppel sought to close this major gap by proposing a National Assessment of Educational Progress to provide data on the quality of learning in the Nation's schools. There was considerable opposition on grounds that the assessment would lead to federal control of education and a national curriculum. Similar opposition greeted the Elementary and Secondary Education Act, also proposed by Johnson and Keppel, which had as its centerpiece Title I to aid low-income students. That law passed in 1965.

The National Assessment, though, was not launched until 1969. It emerged in a form that assuaged the fears of its critics but severely restricted its public impact and significance.

In recent years, though, the tide of opinion has turned. The U.S. Department of Education was established under President Carter in 1979. In 1983, the National Commission on Excellence in Education, appointed by Education Secretary T. H. Bell, issued its report, "A Nation at Risk." The commission somberly documented "a rising tide of mediocrity" in American schools and summoned a national movement for education reform. Bell also issued the first "wall chart" using data from Scholastic Aptitude Tests (SAT) and the American College Testing (ACT) Program to compare academic achievement in the 50 states.

Meanwhile, statewide testing programs proliferated. Almost all made public district-by-district and school-by-school comparative data. Many set standards of expected performance.

In 1988 NAEP was authorized to conduct voluntary state-by-state assessments in eighth grade math in 1990 and in fourth and eighth grade math and fourth grade reading in 1992. The same legislation created the Governing Board as an independent policy-making body for NAEP and authorized it to improve the "form and use of the assessment and to set "appropriate achievement goals."

During the past year the issue of national education goals has come to the forefront at the Charlottesville Summit of President Bush and the Nation's governors and in subsequent actions by the President and the National Governors' Association.

The need for national goals and standards was stated clearly by the Southern Regional Education Board in its 1988 report, Goals for Education:

> "If excellence means anything at all, it is a universal concept...We must be measured against the same criteria of excellence which are applied everywhere...That bold claim was controversial when made by the Southern Regional Education Board nearly three decades ago...Today, there is wide agreement that SREB states should strive for national standards. And some, particularly governors, assert that international standards are more appropriate now that the marketplace is increasingly global."

As Ernest Boyer, president of the Carnegie Foundation for the Advancement of Teaching, has declared, "The failure to establish understandable criteria and standards (for educational assessment) will lead to loss of confidence and a huge erosion of public support for the Nation's schools. We (must) give the public some evidence that our schools are working and that our $180 billion investment is paying off."

"We are now trying to...develop (national) criteria by which the performance of education can be assessed," Boyer continued, "while at the same time we retain vitality at the local level... If we could get standards straight, then we give schools some yardsticks by which they would be measured, and then we should give them a lot freedom to get there."

Setting appropriate achievement levels on the National Assessment is a step in that direction.

The Need for Appropriate Achievement Levels

For the past 20 years the National Assessment of Educational Progress, like virtually all nationally standardized tests in the United States, has reported results in terms of average performance. Sometimes it has announced what proportion of students knew a certain fact or could demonstrate a certain skill. But it has shied away from saying clearly whether average performance was good enough or whether the facts and competencies it tested were ones that students really ought to know.

Of course, the NAEP assessments, like other tests, implicitly do contain judgments of significance and expected performance. Why test anything unless somebody thinks it's important? In developing NAEP, there has long been an elaborate consensus process, involving teachers, university professors, and interested groups, to determine rather precisely what body of knowledge and skills each test should measure. But again, the tests themselves and the committees creating them have only implicitly provided a basis to say how good is good enough.

As the National Academy of Science said in a report (1982), NAEP "was conceived as a white paper on the status of education in America." Its primary purpose is to report to the public on the quality of learning in the schools. But until now, the significance of its findings has often been unclear.

In an effort to improve reporting, NAEP in recent years has said what proportion of students in different grades reach different proficiency levels, but these levels--200, 250, 300, etc.--have been derived from the distribution of test results themselves, not from any prior judgment of what students ought to know. Each 50 points up or down represents one standard deviation, a measure of variation in test scores. The cluster of skills that differentiates each major level is determined by looking at the patterns of right and wrong answers after the results are in.

While helpful, such proficiency levels, are in truth simply statistical distributions. They provide limited guidance for determining whether students have mastered a challenging curriculum or have acquired the knowledge and skills needed to advance in school or move on successfully to college and adulthood.

Defining what performance ought to be--and providing strong justification for the judgment used in making these definitions will greatly enhance NAEP's central function as a yardstick of educational achievement.

Framework and Definitions

The Committee recommends that the Governing Board adopt a framework for setting appropriate achievement levels that includes three levels of achievement for each grade and subject on NAEP.

The central level will be called **Proficient**. It will represent solid academic performance for each grade tested--4, 8, and 12--and reflect a consensus that students reaching such a level have demonstrated competency over challenging subject matter and are well prepared for the next level of schooling. At grade 12 the proficient level will encompass a body of subject-matter knowledge and analytical skills, of cultural literacy and insight, that all high school graduates should have for democratic citizenship, responsible adulthood, and productive work.

There will be one higher level, called **Advanced**, signifying superior performance beyond proficient grade-level mastery at grades 4, 8, and 12. For 12th grade the advanced level will show readiness for rigorous college courses, advanced technical training, or employment requiring advanced academic achievement. As data become available, it may be based in part on international comparisons of academic achievement and may also be related to Advanced Placement and other college placement exams.

There will be one level below proficient, called **Basic**, denoting partial mastery of the knowledge and skills that are fundamental for proficient work at each grade--4, 8, and 12. For 12th grade this will be higher than minimum competency skills (which normally are taught in elementary and junior high schools) and will cover significant elements of standard high school-level work.

The Board will ensure that the content of each subject-matter assessment supports three achievement levels at each grade with clear distinctions between them. It will encourage research to permit use of international data in defining achievement levels.

This framework, applied through a broad consensus process to specific subjects in the National Assessment, will provide meaningful benchmarks of academic achievement. However, unlike any single measuring point for each grade, it will also show a wide distribution of student performance.

These benchmarks will permit states and the nation to see what proportion of students have reached very high levels of achievement on NAEP exams; strong, acceptable levels; and levels of partial mastery. Thus, it will provide a measure and incentive to improve the learning of all segments of the distribution--bottom, middle, and top.

The framework of three achievement levels at each grade is not a warrant for tracking. Indeed, the NAEP tests and the achievement levels based on them will help to ensure that all students attain competency in challenging subject matter.

The proposed achievement levels will define levels of learning tied to a common core of knowledge and skills that ought to be available to all students, regardless of family income, ethnic background, region, or type of community. The achievement goals on the National Assessment will serve to underscore the point that American schools ought not to water down what they teach the poor and beef up what they offer the more affluent.

Procedures for Establishing Specific Achievement Levels

The process for determining achievement levels should be an outgrowth of the national consensus effort used in developing the content and objectives of National Assessment exams.

For many years NAEP has reflected a broad consensus, regularly updated by representative committees, on what is important for students to learn. In each subject area different topics at different ranges of difficulty are assessed at different grades, reflecting a consensus judgment on curricular emphases and objectives.

The proposed achievement levels will add to assessment frameworks and objectives the specific definitions of basic, proficient, and advanced achievement at each grade tested, which are based on the content of National Assessment exams. These are not broad general goals of education or curriculum, but substantive descriptions of levels of achievement tied firmly to National Assessment questions and objectives.

To assist in setting achievement levels for specific subject areas the Board will appoint ad hoc advisory panels. These will consist of state and local educators, scholars, employers, civic group representatives, and other interested citizens. The panels will be charged with using a proven judgment procedure to recommend which test questions and/or which proportion of questions students need to answer correctly to reach different achievement levels.

As part of this process, the panels will be required to prepare detailed descriptions of the subject-matter knowledge and skills proposed for each achievement level. These definitions will be based on the general descriptions adopted by the Board and will be accompanied by an explanation and rationale for the definitions proposed. It is important that there be a clear distinction between each proposed level.

The definitions of achievement levels will be similar (though presented in more detail) to the descriptions of NAEP proficiency levels prepared since 1985 by Educational Testing Service, the NAEP contractor. But, unlike the previous proficiency levels, the descriptions of achievement levels will be based on an informed, coherent judgment of what students ought to know rather than on the distribution of test results.

In preparing descriptions of achievement levels and assigning test items to them the panels should not only use their own judgment and expertise but should take into account a wide range of background information and frames of reference. These may include relevant curriculum and testing data from state, local, national, and international levels; comments solicited from interested citizens, specialists, and education agencies; research on the performance of different groups, such as literate young adults; or studies equating NAEP to Advanced Placement, Armed Forces, business, and other testing programs.

The advisory panels should refer to at least some of these sources or others in presenting and justifying their proposed definitions of achievement levels.

To illustrate the content of each proposed level, the panels --with staff assistance--will provide representative sample test items, similar to the illustrative items that have regularly been published in NAEP objectives booklets and reports. These will be accompanied by correct answers for multiple-choice items and scoring protocols for any essay or other open-ended questions.

The proposed definitions, illustrated by sample questions, will be submitted to the Board for approval. The Board will seek wide public comment before acting on the panels' recommendations.

Reporting NAEP in Terms of Achievement Levels

After appropriate achievement levels are approved by the Board and the questions and/or proportion of questions that students must answer to attain them are determined, the levels will be placed on the NAEP scoring scales. The proportion of students attaining each level will be reported.

The three achievement levels developed for each grade will be mapped onto an achievement scale. These levels will become the primary means for reporting NAEP results. However, scores at each quartile will also be reported as another means of showing the distribution of performance.

There may be advantages in using separate scales for each of the three grades in NAEP as this may be a more meaningful and educationally significant way to present assessment results. Such scales may show more clearly the variations in performance for each grade and subject in the assessment.

The scale for each grade--with basic, proficient, and advanced achievement levels clearly defined--would be distinct from any subscales for particular skills. It may be distinct from any common cross-grade scales, spanning grades 4, 8, and 12.

Under current practice, initiated six years ago, all NAEP data for each subject, such as reading or mathematics, are reported on a common scale that spans grades 4, 8, and 12. These subject-matter scales have a uniform mean score of 250, based on the performance of students in all three grades tested. Each 50 points represents one standard deviation across all students in all three grades. Because the same scale applies to grades 4, 8, and 12 the variations for each grade and subject tend to be small, especially for grades 4 and 8. For example, with only one common scale for mathematics, almost no 4th grader will ever be at the advanced level even though a sizeable percentage of 4th grade students may be doing what is advanced work for the 4th grade.

Once well-developed achievement levels are established, it is the National Assessment Governing Board's intent that the stability of the achievement levels be maintained over a period of several years, perhaps a decade. Test items may be updated and the test framework may even be changed, but priority will be given to maintaining the stability of the achievement levels.

If the three-achievement level format for reporting is successfully developed, this will provide more detailed information for each grade level. Even though variations in performance within each grade will be shown more clearly, it remains to be determined whether such more detailed information will overcome the perceived shortcomings of NAEP's across-grade scale. The Board will pursue this unanswered question as it relates to the assessments of 1992 and subsequent years on a timeline to be developed by Board staff in consultation with staff of the National Center for Education Statistics and the Educational Testing Service.

When Should Achievement Levels Be Set?

The Committee recommends that the Board adopt the proposed framework and procedures for establishing appropriate achievement levels as policy for all future NAEP assessments. It should begin setting achievement levels with the 1990 assessment of mathematics.

The mathematics assessment is well-suited for setting appropriate achievement levels. It has been thoroughly revised through an extensive consensus process, conducted by the Council of Chief State School Officers, and incorporates many elements recommended by the National

Council of Teachers of Mathematics. The assessment includes a progression of challenging topics that goes well beyond the level of basic skills where NAEP assessments have usually concentrated in the past.

The content and objectives of the math assessment have won wide endorsement from mathematics educators and state education departments. The assessment involves a field where substantial consensus already exists.

If the Board approves this proposal, it should follow the timetable adopted by NAGB on March 2, 1990. The timetable provides for the Board to appoint the panels to recommend specific mathematics achievement levels by mid-September. A public hearing or forum on these recommended levels would be held in mid-October. The Board would take final action on the mathematics achievement levels at its meeting of November 16-17, 1990.

Such a timetable would permit the achievement levels to be used in the first public reporting of nationwide data on the 1990 math assessment during the summer of 1991. State-by-state results would be reported in terms of appropriate achievement levels only at the request of individual states. The states did not know that such achievement levels would be established when they agreed to participate in the assessment. However, many states may be interested in receiving this information at the same time other state-level data are released.

This first effort at setting appropriate achievement levels should be seen as provisional and subject to further refinement and change. However, it is anticipated that the achievement levels defined will remain in place when the mathematics assessment is repeated in 1992 and for several subsequent math assessments. Soon after the math levels are set, the Board may wish to begin planning, based on that experience, to set achievement levels for the 1992 assessments of reading and writing.

NAEP and International Achievement Levels

As the Governing Board declared in December, the National Assessment ought to become a major vehicle for comparing the achievement of American students with those of other countries. International data on student performance should be used in establishing appropriate achievement levels on NAEP exams.

The Committee proposes that the advanced level on NAEP proficiency scales become a standard of "world-class performance." As data become available, the advanced level should be based in part on high levels of performance on international assessments of student achievement.

To do this in a systematic way data would have to be obtained by having representative samples of students in other countries take NAEP assessment items, as the Board proposed in December. Alternatively, some form of equating of NAEP and other tests given internationally would be required. Some international anchoring could begin with data already available from studies conducted by the International Association for the Evaluation of Educational Achievement (IEA).

A special study was conducted in 1988 by Educational Testing Service as the first International Assessment of Mathematics and Science. In this study math and science items from the 1986 NAEP were administered to samples of 13-year-olds (mostly eighth graders) in five countries and six provincial Canadian school systems.

The proposed advisory panels to set achievement levels for math should consider these data in defining the advanced level for 8th graders on the 1990 NAEP math assessment. This might serve as an important prototype for using international data in establishing achievement levels on NAEP exams and will be helpful in determining what similar data should be obtained in the future.

Rejected Alternative Proposals To Use NAEP for Setting Achievement Goals

Two alternative suggestions have been made for setting achievement goals on the National Assessment in contrast to the appropriate achievement levels proposed in this paper. Both have serious drawbacks, as noted below. The proposals, with comment, are as follows:

1. Use the existing NAEP proficiency levels and set targets on them for the proportion of students that should reach different levels.

The fundamental problem with this suggestion is that the proficiency levels are not based on content but on score distributions. They are determined only after the tests are given with 250 as the mean and each 50 points representing one standard deviation. Since the scales change when NAEP tests change, previous results are sometimes recomputed, according to scales developed from the most recent testing.

In 1990 and 1992 ETS plans to give two different versions of the NAEP to two separate national samples in reading, mathematics, and writing. One version, a copy of old tests, will be used for trend data. The second version, much revised in each subject, will be used for the major cross-sectional reports and for the state-by-state assessments in math and reading. For 1994 the NAEP science test is planned to undergo a major revision through the national consensus process.

Targets might be set on the previous NAEP tests, but these would provide no data on individual states. Further, the older tests (those administered prior to 1990) have the additional drawback that much of the material on them is regarded by experts as outdated or inadequate.

Of course, goals might be set on proficiency levels that ETS establishes for the new NAEP exams. But that can't be done until the tests themselves are scored and scaled and the new levels are created. It is only at that point that anyone will know what knowledge and skills are represented by any particular level and how any level might relate to grade-level learning in school.

At that point, of course, we will know the proportion of students at each proficiency level. Any goal-setting effort would be empty unless it is for the next administration of the test, which will delay the whole process several years more.

There are three more problems with this alternative:

(a) For each subject there are only four or five defined proficiency levels, spanning all three grades tested--4, 8, and 12. This may well be too few for meaningful reporting and to show a distribution of performance at each grade. By contrast, the Committee has proposed nine levels over the same three grades.

(b) As previous data published by NAEP indicate, some of these levels have very little fit with material commonly taught at particular grade levels. Thus, they can say very little about what students have learned.

(c) Choosing what percentage of students ought to perform at a particular level is an arbitrary, poorly-defined exercise. If 5 percent of students are at a certain high level now, should 10 percent reach there in the year 2000? or 8 percent? or 12 percent? or 20 percent? Why??

We believe there is no reasonable basis for the Governing Board to set such targets. Also, there is no statutory warrant for it to try or to attempt to devise a process for doing so.

Setting targets for performance by stating what percentage of students should reach different levels is essentially a judgment that ought to be made by educational and public officials. Defining levels of performance that may serve as appropriate achievement goals on NAEP is a proper activity for NAEP's Governing Board. Others may then use the levels NAGB defines as part of their own goal-setting activities.

2. Report scores by quartiles and set targets for score increases at each quartile point.

This proposal would encounter the same problems in target-setting as the one above.

There is no clear basis for setting such targets and NAGB has no warrant and no particular competence to do so. There is the further problem that no targets would be meaningful unless they were for a test that has been used in the past; both the reading and mathematics tests for the 1990 and 1992 state-by-state assessments are new, vastly different (and we think better) exams, which may not equate to previous National Assessments. The science exam may undergo major change for 1994.

Also, the point values that might be reported for each quartile have very little meaning in themselves and little significance to the public. There simply is no clear definition of the meaning of 265.8--the point value of the bottom quartile for 17-year-olds in the 1988 NAEP reading assessment. If the quartile score went up to 270, that would say virtually nothing about what additional skills or knowledge students might have. By contrast, achievement levels can be defined clearly in terms of what students know and are able to do.

Reporting by quartiles certainly is valuable for making comparisons among groups, showing the distribution of performance, and charting trends. It should continue to be part of the regular NAEP reports and should be given more prominence than it has had in NAEP reports of the past, which often have focused on averages. However, achievement levels are a much more meaningful measure for understanding the National Assessment; these should become the principal means for reporting NAEP results.

Another Suggestion. It has also been suggested that NAGB not set any achievement goals or targets, but rather should devise a process that others might use to set targets for increasing the proportion of students at high levels on NAEP exams.

As discussed under alternative one above, there is no method for setting such targets which is not fundamentally an exercise in estimation and exhortation.

Endnote: The Promise and Some Cautions

Setting appropriate achievement levels on the National Assessment will help define important outcomes of education, stating clearly what students should know and be able to do at key grades in school. This will make the Assessment far more useful to parents and policymakers as a measure of performance of American education and perhaps as an inducement to higher achievement.

As the National Commission on Excellence in Education noted in 1983, it is the nation that is "at risk," not just a few states. It is the whole country that is competing against the nations of Europe and Asia that today are challenging our economic position. In a Gallup poll last September over 70 percent of Americans said they favored "national achievement standards and goals."

Certainly, the Governing Board has no power of command over schools, nor does it seek such authority. NAEP hires no teachers, selects no textbooks, assigns no homework, determines no course requirements, and awards no diplomas. These are decisions made locally and by the states. The states and local governments retain full authority over what is taught in their schools. Even participation in NAEP is completely voluntary and should remain so.

However, by setting appropriate achievement levels through a broad consensus process the Governing Board has an opportunity to define a common core of learning that is important for all American children to acquire. The achievement levels will be benchmarks, points for judgment and encouragement, not edicts or commands.

If they are set well, the achievement levels will increase greatly the significance and meaning of NAEP results. Any further impact they may have will be through a process of persuasion and voluntary acceptance.

Part 2

Technical Procedures

Introduction

The technology for setting achievement levels[3] has been developing over the past 35 years, and is now considered standard operating procedures for many assessment programs at the state and district level.

The technology for setting achievement levels falls into two broad categories: judgmental and empirical. Judgment methods employ appropriate groups of judges to rate the individual items in an assessment on specific criteria related to examinees' mastery or non-mastery of the content. Empirical methods use data collected from various examinee populations to make decisions about cutting scores which discriminate between two or more proficiency levels in the population. The Contrasting Groups procedure is an example of this methodology. In this approach, data from two examinee groups who clearly differ in their achievement level on the assessment are used, and the cut score is placed to maximize the discrimination between these two groups.

Judgment methods can be implemented prior to test administration, since only the items and not item data are required. However, it is highly recommended that item data, including, but not limited to, item characteristic data and distractor analysis, be made available to the panels. It is argued that allowing judges to reconsider their initial ratings and to modify those judgments generally produces more reasonable achievement levels, and reduces variability in the estimates. Item data for the 1990 mathematics assessment would be available in the late summer, and should be used by the panels in this case.

---

[3] In this section of the staff paper the term **achievement levels** continues to be used in order to be consistent with Part 1, even though the literature has typically discussed this methodology in other terms such as standards or performance standards.

Empirical methods require that a trial assessment be administered before setting the achievement levels. It is recommended that empirical validation procedures be mounted subsequent to establishing achievement levels. Validity studies are essential in order for the achievement levels to withstand the scrutiny of the educational, business, and public sectors. It is also recommended that external validation studies be conducted where NAGB could compare the classification of groups of students according to the NAEP levels with their classification by a variety of external criteria. At the fourth and eighth grade the criteria would be school-related, whereas, at the twelfth grade criteria should include school-based and post-graduation outcome measures.

A Modified Angoff Procedure

While there are a number of competing judgment procedures that could be used for setting achievement levels, often times yielding different results, a modified Angoff procedure is recommended for a number of reasons. First, the advantages and disadvantages of many of the competing procedures are well documented in the literature. There have been any number of research studies completed documenting some of the differences; the Angoff procedure is generally superior. Secondly, it is quite straightforward; both the judging task and its results are intuitively interpretable. Thirdly, it does not require the administration of items to a trial population. This means, of course, that setting achievement levels can begin immediately. However, since item data will be available, it should be used by the panels in this case. For all these reasons, and perhaps others not mentioned here, the Angoff methodology is clearly the methodology of choice.

The Angoff method will be modified to accommodate the fact that NAEP is not attempting to define the probability of a "minimally competent" student getting an item correct. As described in an earlier section of this paper, NAGB is defining achievement levels at three benchmarks on the scale, basic, proficient, and advanced.

Assessment Content

A national consensus process is used to arrive at the content objectives of each subject assessed. The specific details of the process varies from subject to subject. However, the overall concept involves various publics in advising the Board on the current theoretical, curricula, and instructional status of any given content area. The process includes numerous iterations filtering each perspective through that of competing ones, until a final product is derived which represents the best thinking in the field and for which there is general agreement.

In the basic areas, such as reading and mathematics, and, indeed, in all the NAEP core areas, there is an underlying assumption of a developmental curriculum. That is, specific objectives span several years as the students' capacities develop from the lower levels of the content taxonomy in the elementary grades to the highest levels at the upper grades. This approach ultimately forms the conceptual basis of the NAEP scales which currently cut across grade levels and are behaviorally anchored to real tasks and accomplishments at specific intervals on the scale. The content objectives are then defined in measurable terms as the consensus process continues to spell out the test and item specifications. In other words, the consensus process moves toward articulating not only content expectations at each grade level, but the parameters within which those objectives will be assessed. Typically, the field testing of an item pool follows and the final selection of appropriate assessment items is made by the Board.

Achievement Levels

In identifying the content specifications for each subject area assessed, there is an underlying assumption that all students in grade 4, for example, should be able to respond to questions about the "volume of rectangular solids." In other words, this objective would not have been assigned to grade 4 if the framework had not placed it there. This is a reflection of the criterion-referenced nature of NAEP. However, due to measurement error in the assessment, and due to the less-than-perfect performance of students on the assessment, in any given grade level there will be a distribution of performance. So, even though the "ideal" expectation for grade 4 as described by the test objectives might include knowledge of the "volume of rectangular solids," a more accurate expectation for grade 4 can be derived by the careful examination of the items designed to measure the grade 4 assessment objectives.

Achieving consensus on the real expectation for students is the process of setting achievement levels, the yardstick by which the degree of success on the subject matter content for each grade will be assessed.

Setting definitive achievement levels for each grade and in each subject area assessed allows users of NAEP to make informed judgments about the quality of the results, and seeks to provide answers to the following questions: How good is good enough? Do we have substantially different expectations for different content areas? Are there levels of achievement within each content area that distinguish those who are truly proficient in the content from those who are only modestly proficient? Setting achievement levels for NAEP will assist us in answering those questions, and in interpreting the data better.

Number of Levels and Scales for Each Grade

Earlier it was mentioned that three achievement levels would be established for each grade level. We must caution, that in order to accomplish three levels at each grade level, the distribution of item difficulty and content must be adequate (1) to support the accurate and precise description of collective examinee performance in the four achievement regions defined by the achievement levels, and (2) to describe examinees' collective abilities to perform tasks that are deemed to be clear and interpretable by educators and the public.

At the present time, with a single cross-age/grade scale, there are five benchmarks. If three unique grade scales are established, with three benchmarks each, this results in nine achievement levels, four more than NAEP now has. It is not clear at this point whether or not the data will support this increase. However, preliminary judgments seem to indicate that it should. This issue certainly will need to be reexamined for each subject area, particularly as the one hour response time for examinees is used to provide more extended responses on fewer numbers of items.

On how many scales or subscales should achievement levels be set? A sufficient number of scales should be created to represent accurately achievement on all or nearly all of the exercises in the pool at a given grade level. As many exercises as possible should be incorporated into the IRT scales. This may entail some revision of initial plans for scaling. It must be recognized, however, that small, important groups of exercises may remain, which are insufficient to support separate IRT scales but sufficiently important and substantive enough to warrant not setting aside. In such cases, item clusters may be scaled using alternate techniques. Scale scores developed by alternate methods should be expressed in metrics comparable to those used for IRT-based scales.

When more than one scale is required to represent accurately achievement on all or nearly all of the exercises, an index should be created by taking a weighted composite of scales, the weights to be determined by a rational, deliberative procedure. Whenever possible, achievement levels should be established and reported for all scales as well as the composite indices.

Procedures for Setting Achievement Levels

There are probably hundreds of variations on what has become known as the "Angoff Method." This is because a method for setting achievement levels includes much more than simply the nature of the judges' rating task. In developing the method to be implemented, reference and consideration must be given to the following features of the process discussed here.

Composition of the Panels. The groups to be represented on the panels must be identified, and procedures for selecting representatives must be determined. It is recommended that the panels be composed of individuals with expertise in the education of students of the ages and grades under consideration, in the subject areas under consideration, with experience in the assessment of students' achievement in the subject areas under consideration, with knowledge of the typical subject area achievement of students of the ages and grades under consideration, and, in the case of twelfth grade assessments, with knowledge of the subject area achievement requirements of high school graduates who aspire to post-high school experiences in the work force, the military, or post-secondary education programs.

Major national organizations will be contacted to recommend from among their members individuals who might serve on the panels as well as alternates. In selecting members for the panels great care will be exercised in making certain that the required and desired demographic and technical characteristics are represented on the panels.

There are two additional criteria which must be applied when designing the composition of the panels. First, there should be some continuity with the mathematics consensus panels convened in 1988 to recommend the content and objectives of the 1990 assessment. Therefore, some members of the previous panels should be requested to serve on the panels. The second criteria must ensure that states participating in the 1990 state-by-state trial assessment be represented on the panels as well. This is particularly important at the eighth grade level.

Size of the Panels. How many judges should there be? This is a technical issue which is not easy to answer. Generally speaking, the larger the sample of judges on the panels the less error of estimation there will be. However, every estimation procedure which employs a sample to estimate a population parameter will have some amount of error associated with it. In addition, every instrument has a margin of error associated with it called the standard error of measurement. Setting standards, therefore, does add a second source of error. It is desirable to keep this additional source of error at a minimum, so that the overall standard error is not excessively large.

It is recommended that a sufficient number of judges be on the grade level panels such that the overall standard error is increased by no more than 12%. This can be achieved by ensuring that the standard error of the mean recommended grade level achievement levels is no more than 0.5 of the standard error of measurement of the assessment. The research has suggested that this criterion will probably necessitate having between 16 and 20 judges on each grade level panel, that can be divided into four groups of 4 or 5 judges each. Each group will be chosen, if possible, to be representative of the entire group. In that way, independent replications of setting the achievement levels process can be conducted and the resulting achievement levels compared.

Training of the Judges. It is recommended that training for the panels include training both to the task and the process. This training would include, but not be limited to, definitions of the three achievement levels, the rating method to be used, and the adjudication of extreme ratings through panel iterations. It is critical that the training include practice exercises with feedback, and several simulations to ensure full comprehension of the task, and full understanding of the definitions of the benchmarks. Of special interest will be training judges to provide multiple ratings for each item corresponding to the benchmark points of interest.

Resources Available to Judges. As discussed earlier it is highly desirable to have item characteristic data available to the judges after they have made their initial ratings of items. Allowing the panels to have the data to condition their final judgments usually leads to more reasonable and converging achievement levels. An informed panel is more apt to make sound judgments than an uniformed panel. Since in math the 1990 data will be available at or around the time the panels meet, it is in the best interest of defensible achievement levels that the panels be given such data.

In addition, judges will have the test and item specifications available, the content area framework, and all the items coded by grade and objective, and an answer key.

Briefing materials will also be prepared for the judges that will assist the panels in making a more informed judgment about the objectives and exercises in the assessment. These materials might include, but would not be limited to, a variety of supplementary documents and external criteria that could assist the judges in evaluating their individual estimates of achievement levels in each assessment.

General Meeting Strategies. Each panel member will review the framework of the assessment as well as the test and item specifications. Each judge will then be instructed in how to use the Task Review Form (or a form similar to the one shown in Appendix A). Each judge will complete the Task Review Form, and then, as a group, they will determine a consensus average percent for each objective. In reaching a consensus, the discussion will focus on outlier ratings, and each judge will have the opportunity to reconsider h/er own ratings. This procedure will be completed three times, once for each of the three benchmarks. A final listing of ratings for each objective will be compiled, each representing a profile of the content that a group of students who meet the benchmark criteria should have mastered. These consensus ratings will be added to the Item Review Forms (or a form similar to the one shown in Appendix B).

Once the panels have had the opportunity to work with several practices exercises (items), the judges will complete the item reviews individually. Within the smaller groups of 4-5, judges will discuss their individual ratings to reach consensus. Individual judges will aggregate their own ratings to produce an individual achievement levels, and finally aggregate them to produce group achievement levels. This will be completed three times, once for each benchmark.

The smaller groups of judges will then come together to compare their group achievement levels, and to reach consensus as a panel on a single achievement level, one for each benchmark. It is at this point that empirical data from the assessment will be made available to the panels for their consideration. Should judges wish to modify their ratings before reaching a final judgment they can do so at this time.

Describing the Anchor Points. Once the panels have completed their work, the final ratings of the judges will be aligned with the items on the assessment placed in order of their scale values. This graphic representation[4] will display the location of the items on the IRT scale (if available), the degree of agreement among the panel members, and will be used by the panels to generate the content descriptions of the anchor points. Such descriptions will be accompanied by representative items for each point either from the released item pool or other items written specifically to demonstrate the content.

---

[4] The suggestion for a graphic display was made by Edward Haertel, Stanford University, at a meeting held in Chicago on February 24, 1990, with NAGB and ETS staff.

Documenting and Evaluating the Process. A complete record of the meetings and the process used by the panels will be made, so that problems, inconsistencies, or other issues can be addressed in subsequent achievement level activities.

The Board will conduct a formal evaluation of the process. The evaluation will cover all aspects of the process, from both a technical and policy perspective, and will make recommendations for improving future activities in this area.

Appendix A

Task Review Form

Task Review Form

Strategy: This form should be used with the group of judges to help the group reach a joint understanding of what minimum competency is for each task or objective. (In the form, the word "Task" is substituted for "Sub-Responsibility" for convenience.)

Each judge should determine the percent of times that a task or objective is to be accomplished with no or only a few minor errors. As a group, the judges should reach a compromise rating among their collective ratings.

Form:

Directions: Read each task in the role of delineation statement (domain specification or objective) and determine the percent of times each task (objective) must be accomplished with no or only a few minor errors. For example, consider the following task:

Complete a standard order form for ordering office supplies

For this example, what percent of items that an order form is to be completed must the form be completed with no or only a few minor errors?

Task X. ___________ %

The response is _____ % of the times the order form must be completed with no or only a few minor errors.

Now, ask judges to look at the tasks in the role of delineation profile.

What percent of times should each task be performed with no or only a few minor errors?

Write a percent in the space provided.

1. _____ %
2. _____ %
3. _____ %
4. _____ %
5. _____ %
6. _____ %
7. _____ %
8. _____ %
9. _____ %
10._____ %
11. _____ %
12. _____ %
13. _____ %
14. _____ %
15. _____ %
16. _____ %
17. _____ %
18. _____ %
19. _____ %
20. _____ %
21. _____ %
22. _____ %
23. _____ %
24. _____ %
25. _____ %
26. _____ %
27. _____ %
28. _____ %
29. _____ %
30. _____ %
31. _____ %
32. _____ %
33. _____ %
34. _____ %
35. _____ %
36. _____ %
37. _____ %
38. _____ %
39. _____ %
40. _____ %

Appendix B

Angoff Item Review Form

(Method A)

Angoff Item Review Form

Reviewer's Name: ____

Date:__________

Task (Objective Statement: (insert the task objective number here)

This task objective must be performed _____ % of the time with no or only a few errors.

I. Ask judges to think of a group of persons who are just able to meet this required level of performance for this task (objective). The exam items below were prepared to measure this task (objective). What percent of the group of people that you are thinking about will be able to answer each exam item correctly? Write the percent (between 0 and 100) for each exam item in the column labelled "Initial Percent."

Test Item	Initial Percent	Revised Percent
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %
_____ %	_____ %	_____ %

II. When the judges in the work group have provided their initial ratings, ask them to compare their percents on an item-by-item basis. Also, review the scoring key. Identify the judges who have the highest and lowest percent for each exam item. If they are greatly different (about 20% points difference(then they should discuss why the percents were chosen. They do not have to reach a compromise. Only reconsider their own ratings when there are large differences. If they want to change their percents for any exam item, they should write a new percent in the Revised Percent column.

Part 3

Displaying NAEP Results in Terms of Achievement Levels

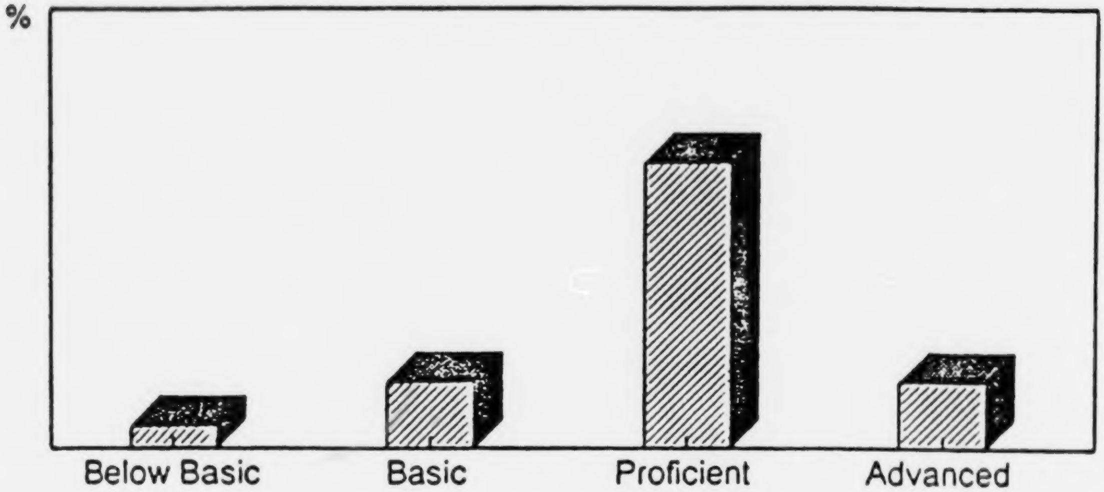
Once achievement levels have been established for a given subject area assessment, the results can be reported in terms of these levels in a variety of ways. Reports of NAEP results can be tailored to specific audiences, thereby increasing the significance and usefulness of NAEP data to educators, policymakers, and the general public.

The graphics on the following pages depict some of the many forms and formats for reporting NAEP results based on the achievement levels. The figures in Sample 1 illustrate two ways to look at performance for the distribution. For a single year, the percentage at each achievement level could be graphed as shown in the first chart. Similarly, the second chart shows changes in the percentage of students at each level over time on successive administrations of a subject area assessment.

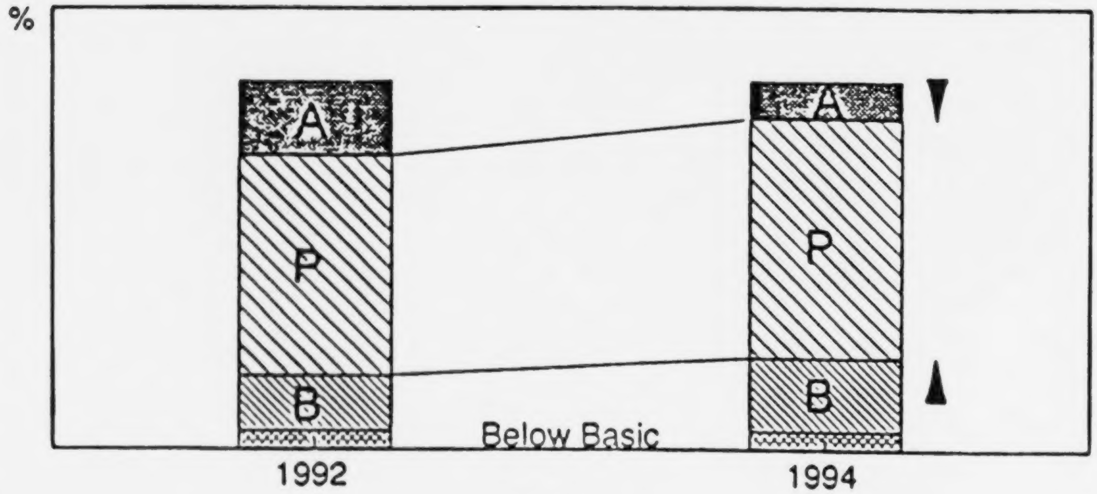
Individual states may wish to set targets by establishing, for example, the percentage of students expected to reach each achievement level. Progress toward these targets could then be displayed, as shown in Sample 2. A value-added approach, as depicted in Sample 3, could present the progress toward a state-defined goal over time. Finally, Sample 4 illustrates the use of achievement levels to show gaps between various subgroups on the NAEP scale.

These charts, though general in nature, do serve to illustrate some of the many ways in which the NAEP achievement levels can enhance the interpretability and usefulness of the National Assessment results for diverse audiences.

PERFORMANCE FOR THE DISTRIBUTION



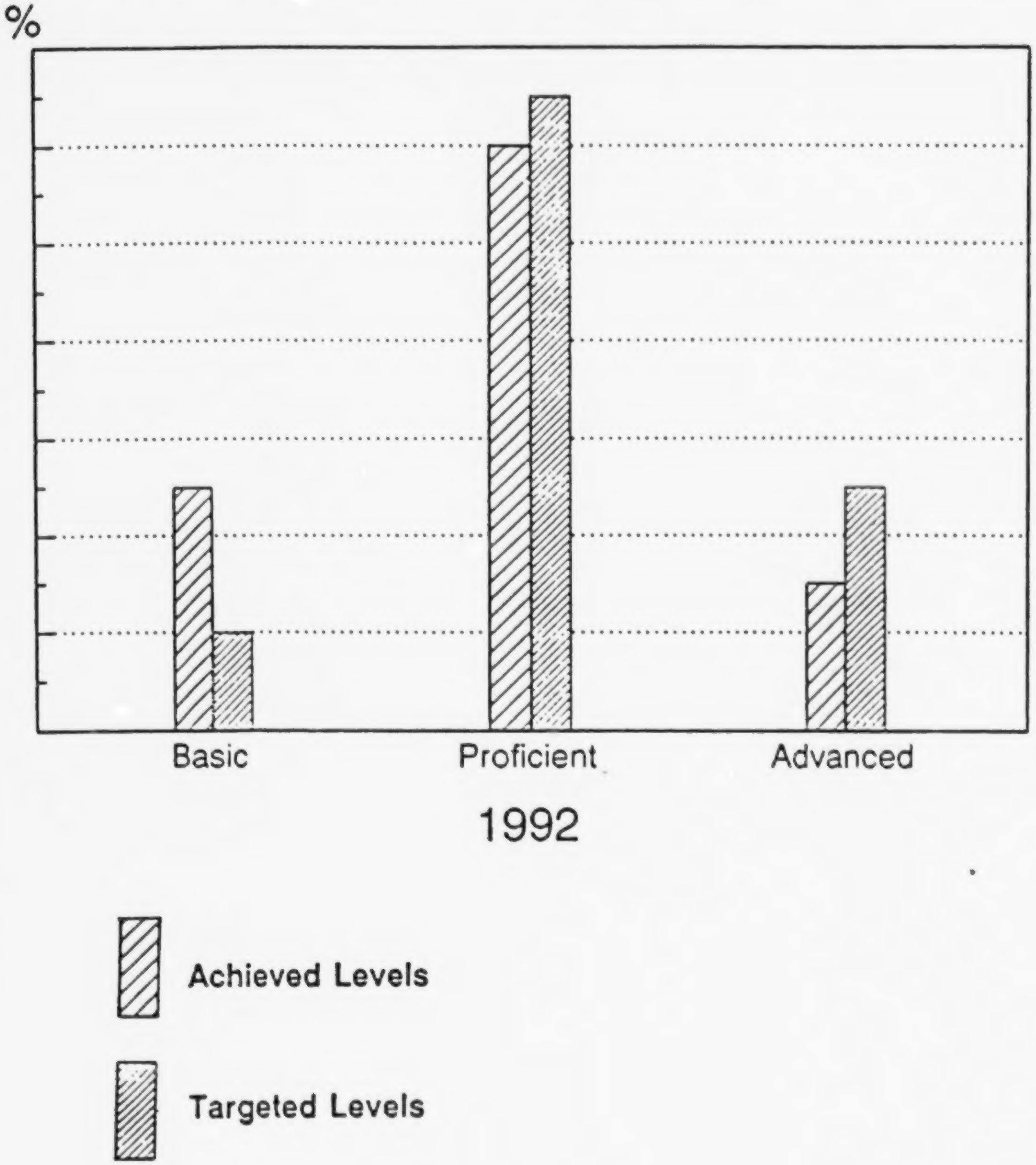
Percentage at Each Level
1992



Percentage Change at Each Level Over Time

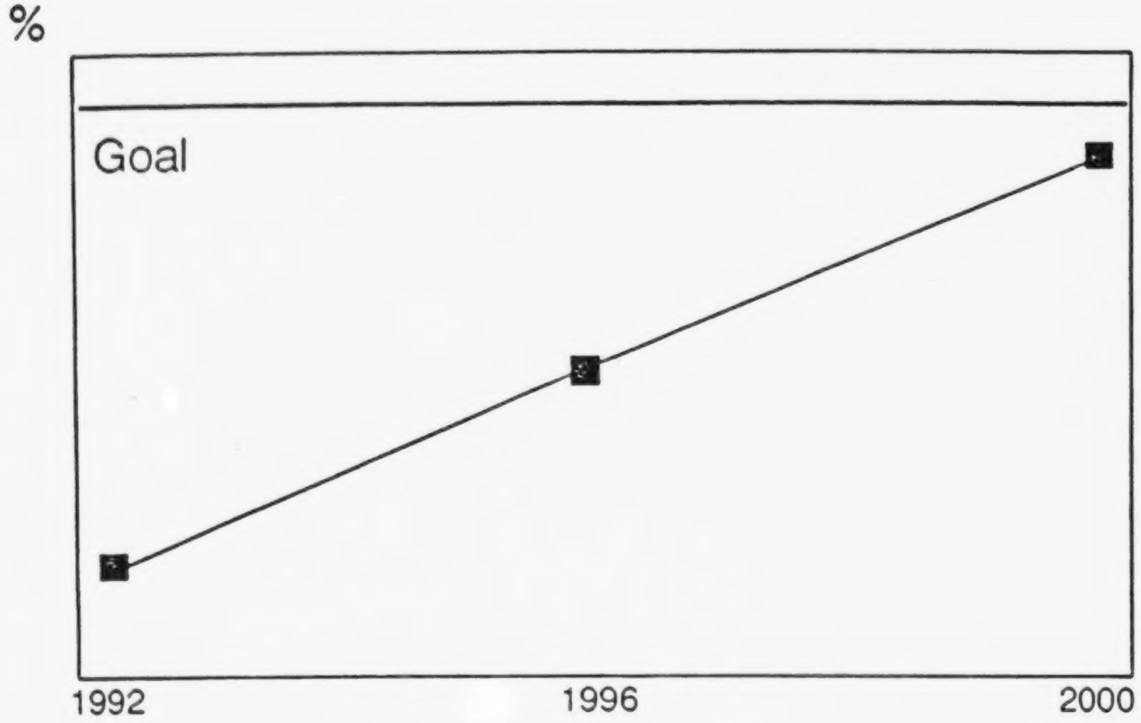
SAMPLE 2

Progress Toward Targets



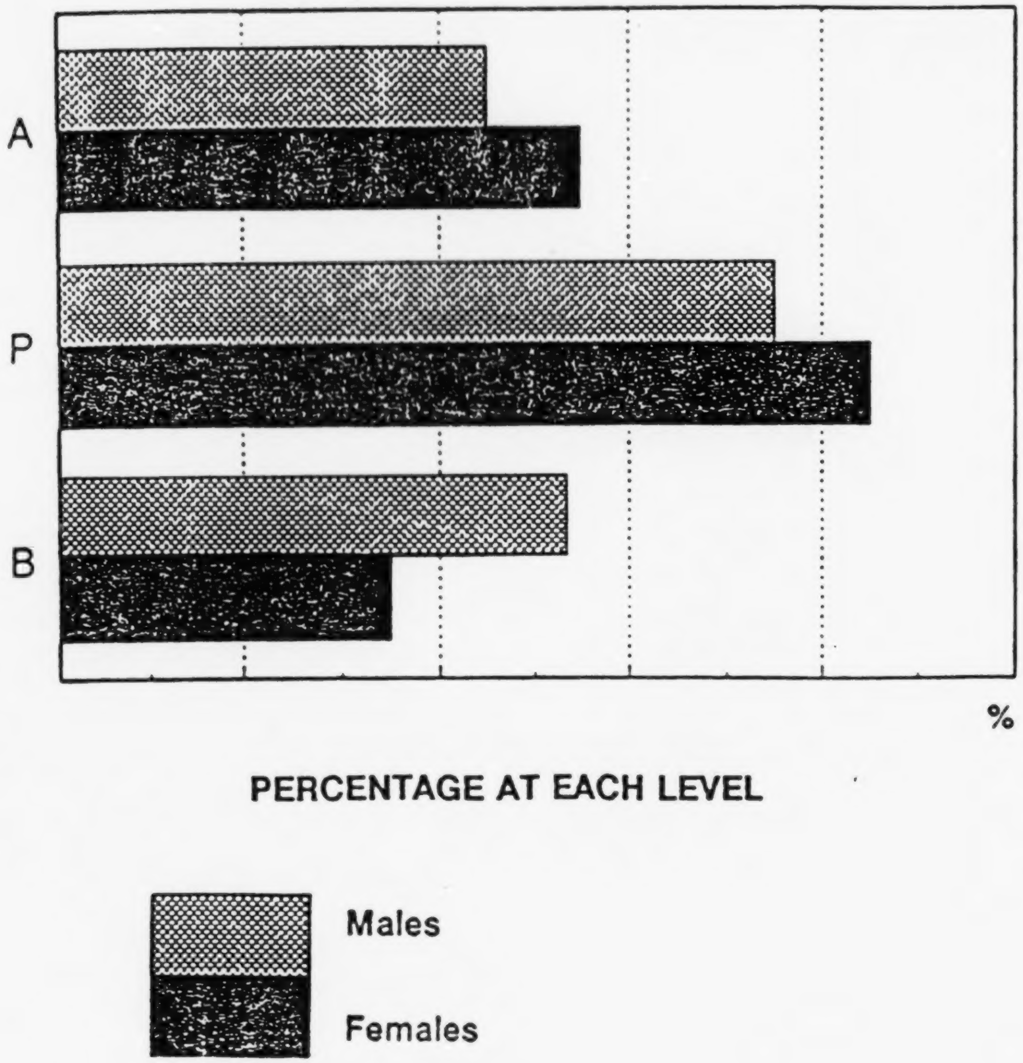
SAMPLE 3

Growth Over Time - Value Added Approach



SAMPLE 4

Gaps Between Subgroups



Appendix K

Replication/Validation Plan

Setting Achievement Levels on the
1990 Mathematics Assessment:
A Validation Plan

March, 1991

Setting Achievement Levels on the
1990 Mathematics Assessment:
A Validation Plan

Introduction

More than a year ago the National Assessment Governing Board began an initiative to set achievement levels for the National Assessment of Education Progress. This task is not only challenging, but is unprecedented in the twenty-year history of the National Assessment of Educational Progress. Performance of American students on the National Assessment of Educational Progress has always been reported in terms of what students know and do in a particular subject area such as mathematics. If achievement levels are established for the National Assessment of Educational Progress subject areas, the nation can know not only what students know and can do, but would also have an important judgment about what students should know and should be able to do. In short, the National Assessment achievement levels will be performance standards that answer the question, "How good is good enough?"

The first step to develop achievement levels began with mathematics which was assessed at grades four, eight, and twelve in 1990, including a trial state assessment at grade eight. Thirty-seven states have assessed the mathematics ability of their eighth graders and will receive individual state reports on that performance in June of 1991.

The process for setting achievement levels is an ongoing one that will span much of the first half of the 1990s. The work on the first effort to set achievement levels in mathematics has shown both the importance and the complexity of the task. After more than a year, additional work is still required before the Board will reach a decision regarding the 1990 mathematics achievement levels. The decision on the 1990 achievement levels in mathematics will likely be reviewed in light of what is learned in this first phase of the process and either confirmed or revised for reporting on mathematics achievement in the 1992 National Assessment. Enough work has been completed to date on the initial effort to set mathematics achievement levels to allow individuals and groups to comment on both the process and the progress. Several extensive evaluations and/or secondary analyses have been completed that contribute to a fuller understanding of the proposed levels and that provide both technical and policy commentary on the levels and how they were derived. These commentaries have raised issues about the levels that need to be addressed as the Board moves ahead with its plan to report the 1990 NAEP mathematics results and to develop achievement levels for 1992 and beyond.

The Board, therefore, consistent with its role as the policy-making body for NAEP, and taking the advice of many thoughtful groups and individuals, has decided to conduct a validation study of the achievement levels before reaching any final decision. The validation process will consist of a series of activities designed to provide evidence of validity for the achievement levels. The five major components of the process are described below. It should be understood that these activities are not developed at this point in great detail. However, it is felt that these five tasks will, if completed in a timely manner, provide the Board with critical validation evidence to assist them in reaching a final decision.

The plan described here was approved on February 12, 1991 by the two Board committees charged with the responsibility of monitoring the achievement levels process. The following briefly describes each task of the plan with an approximate timeline.

Validation Plan

Task 1: Technical Report

It was mentioned earlier that the Board undertook this initiative over 14 months ago. During this period many aspects of the project have been completed. Materials were produced for meetings, documents developed as a result of meetings, and many individuals and groups involved. While this documentation exists, it has not been systematically collected and presented in the form of a technical report. This is required if the process is to be understood and accepted.

Therefore, a comprehensive technical report will be prepared as part of the validation that will address the technical aspects of the process as well as the Board policies implemented through various technical decisions. The report will be prepared by Drs. Ronald Hambleton, principal consultant, and Mary Lyn Bourque, NAGB staff, and will be reviewed by the Technical Advisory Committee on Standard Setting (TACSS), as well as by selected user-groups such as the state testing coordinators and others. A table of contents and the list of appendices will be prepared in the next few weeks so that work can begin on this important and critical task as soon as possible.

Task 2: Executive Summary

As important as the technical report may be, a shorter, less technical summary is also a critical aspect of validation. The work of the Board and the product they are considering must be accessible, understandable, and useful to a wide audience of stakeholders, interest groups, and publics, including legislators, federal, state, and local policymakers, the business and industrial communities, and most especially teachers, parents, and students. Therefore, a short, focused summary of the achievement levels process, including the next steps to be taken in the validation process, will be prepared to respond to the needs of this larger audience. The report will be prepared by Mr. Larry Feinberg, NAGB staff, and will be reviewed by the Ad-hoc Committee on Validation (ACV), as well as by selected user-groups.

Task 3: Site Validations

The centerpiece of the validation effort will consist of four (4) regional/state meetings designed to collect structured feedback on the product of the Board's efforts, namely, the proposed achievement levels.

Location. Since NAEP collects data from students representing each region of the country, four meetings will be held in March, one each in the Northeast, South, Midwest, and West. Four state departments of education have already offered to assist the Board in conducting these meetings.

Participants. Approximately forty-eight (48) mathematics teachers and twelve (12) non-educators for a total of sixty (60) participants will be invited to a one-day session in each location. The criteria for teacher participation are: (1) teachers must currently provide **direct instructional services** in mathematics to students in grades 4, 8, or 12, and must represent teachers of students with varying ability levels; (2) as a whole, the regional group must be representative on the basis of gender and ethnicity; (3) as a whole, the regional group must include both novice and experienced teachers, and must be drawn from urban, suburban, and rural communities of varying sizes.

The criteria for selection of non-educators is the same as the criteria that was used to identify participants for the original panel. That is, leaders of business and industry, professional groups, parents, individuals who have shown an interest in education, as well as persons who have initiated or implemented school-business partnerships, are all eligible candidates. Naturally, those selected should contribute to the overall representativeness of the group in terms of gender and ethnicity. The state department representative will assist in identifying teachers and non-educators in their state/region who collectively will meet these criteria.

Activities. The one-day session will include a modified training activity for participants, an **independent** rating of a sample of items, an opportunity for participants to judge the proposed achievement levels against their own ratings, and to comment on the proposed cut scores, descriptions, and sample items. Written, structured feedback will be solicited from each participant with no attempt to reach consensus. This information will be synthesized for the Board and presented in such a way that the Board can consider it when making the final decision.

A scripted video tape will be prepared so that all four presentations will be standardized, and participants will not be biased by the presenter in their approach to the task. This approach also ensures consistency in training and group preparation. The tape could be divided into three segments: (1) initial training and preparation of the group; (2) calculating of ratings and comparison of these ratings with proposed cut scores; and (3) collection of structured feedback. The tape will systematically lead the group through the packet of materials distributed at the meeting. The NAGB staff person at each site would be responsible for coordinating the meeting, ensuring a standardized approach, and answering questions that the participants might have.

All procedures will be field tested locally before any meetings are conducted so that the scripts can be refined and finalized, and timing of the tasks (which was such a problem in earlier meetings) can be properly scheduled.

Each participant will be asked to provide **one set** of ratings for a marginally BASIC, PROFICIENT, and ADVANCED group of students on a sample of items. Since item samples are already part of the NAEP BIB spiral design, actual NAEP item booklets will be used by the participants. They will also have the appropriate manipulables such as calculators, protractors, and rulers. If approximately 50 participants rate one of seven booklets at each grade level, that will yield about 5 ratings per item per region, or 20 ratings per item across all four meetings. This arrangement also meets the need for ensuring better item security by not divulging the entire item pool to each participant, and is not unlike the procedures used by the Department in conducting item reviews.

After providing an independent rating of the item samples, each participant will be instructed in how to estimate their sample cut score. They will also be given the cut scores of the original panel and other relevant data and then asked to critique the cut scores in the light of their own professional judgment. In addition, participants will be asked to provide commentary on the proposed descriptions and the sample items associated with the levels. This commentary will be collected using feedback protocols specifically structured to probe the issues (e.g., whether there is sufficient justification for an ADVANCED level given the content of the assessment).

Subsequently, the data collected through this validation process will be analyzed and made available in the Technical Report and other documents related to the achievement levels process to better inform any future endeavors in this area.

Task 4: Final Review by Math Panel

The subgroup of the original 63-member Vermont panel will be reconvened to review the data collected in the validation effort. If the results of the validation produce achievement levels that are substantially the same as those currently being recommended, then there may be only a need for modest revisions. Alternately, if the results of the validation produce results that are significantly different from those produced in the original process, the work of this subgroup will be to develop some recommended options from which the Board can make its final decisions.

Task 5: Response to Evaluations

While the Technical Report and Executive Summary will no doubt address many of the issues raised through the Stufflebeam evaluation, the Technical Review Panel's secondary analyses, or the National Academy's State Trial Assessment evaluation, there is no mechanism for correcting factual errors, or for presenting competing explanations of the data. A formal rejoinder is required to "set the record straight," and to present alternative hypotheses.

Ron Hambleton has expressed an interest in following up on this. It may require some additional analyses, perhaps even some additional information from the panel. However, responding to criticisms in a reasoned way and from a data-based posture is an essential aspect of the validation process. Tasks 1, 2, and 3 alone will not answer all the questions raised in these documents. Task 5 is critical since this is a trial program, and debate and discussions of both the methods of standard setting and the results is important for technical and policy reasons.

Summary

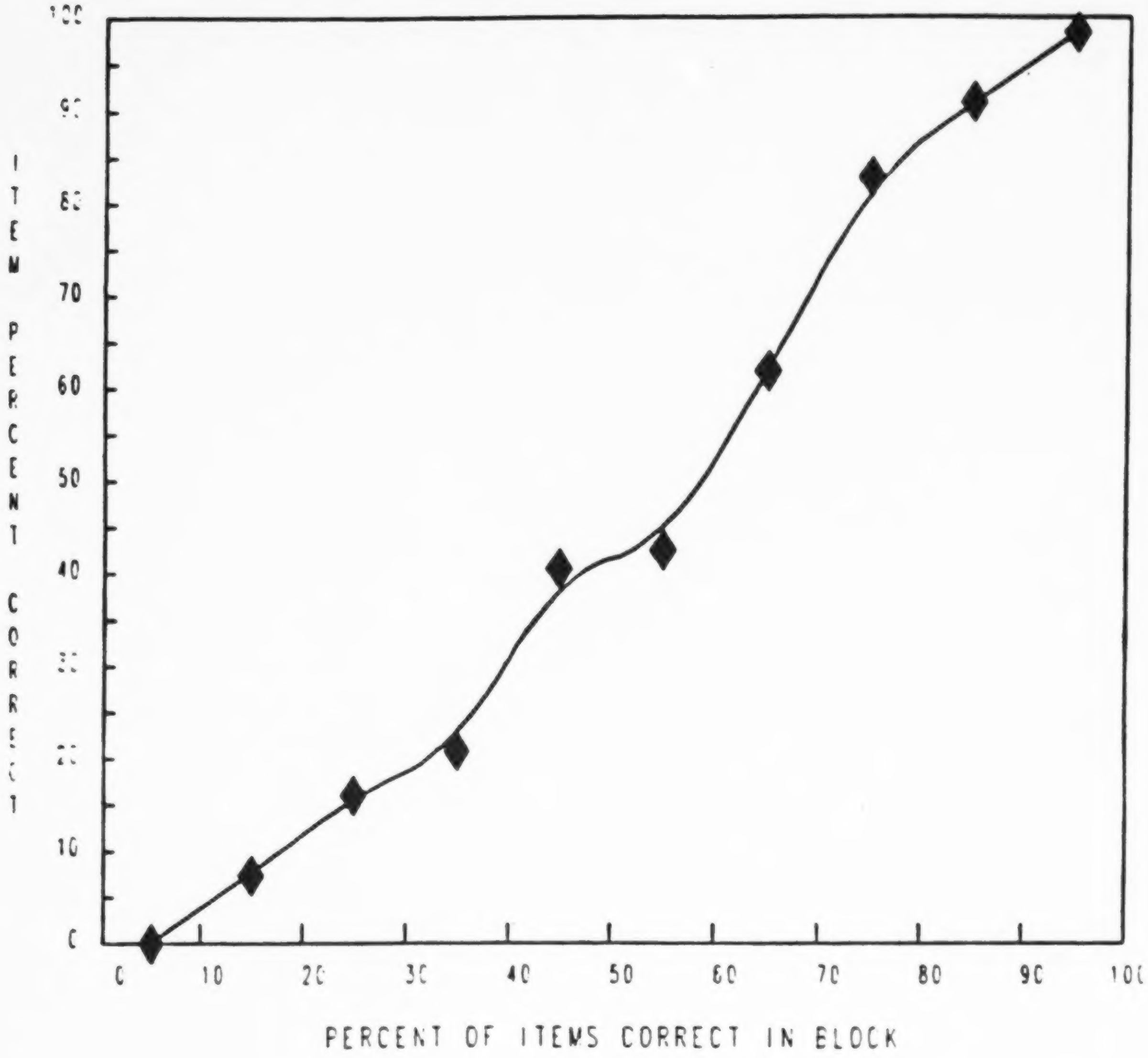
The Board will use all the information and feedback produced in the achievement levels process, the initial recommendations of the original panel, the results of the validation activities, and the final recommendations of the subgroup of the math panel, to make their decision on the achievement level setting effort, and to decide whether to use the levels for reporting the results of the 1990 NAEP mathematics assessment.

Postscript. While the procedures outlined here may appear at first glance to be a short-term process, the work of validation is a continuing one which will proceed well beyond the tasks described. For example, one of the Board's initial goals in exploring achievement levels as a reporting mechanism was to "improve the form and use of NAEP results." Therefore, if the results of the 1990 mathematics assessment are reported in terms of the achievement levels, it would be advisable for the Board to gather evidence on the utility of the levels to users of NAEP data. The utility and understandability for policymakers, which can only be obtained after the results are released in June, is an important component of determining the intrinsic value of setting standards on any assessment, especially NAEP.

Appendix L

Sample Trace Lines
and
Actual ICCs Used in Phase 1

ITEM PERCENT CORRECT BY BLOCK SCORE
GRADE 8 MATH: BLOCK ME



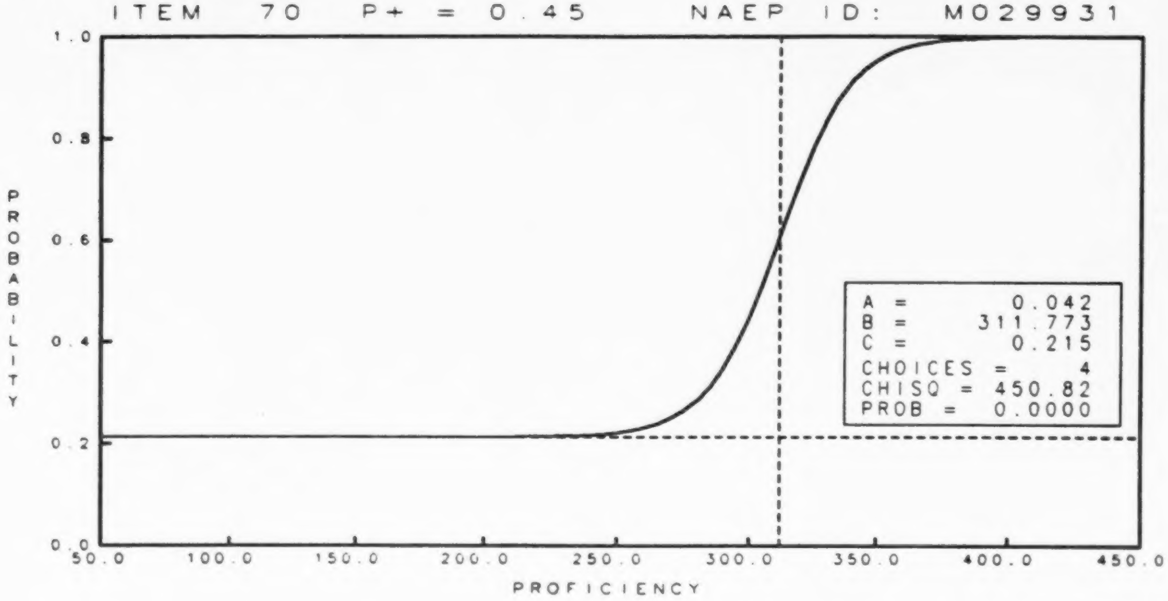
OVERALL ITEM PERCENT CORRECT 0 4606

429

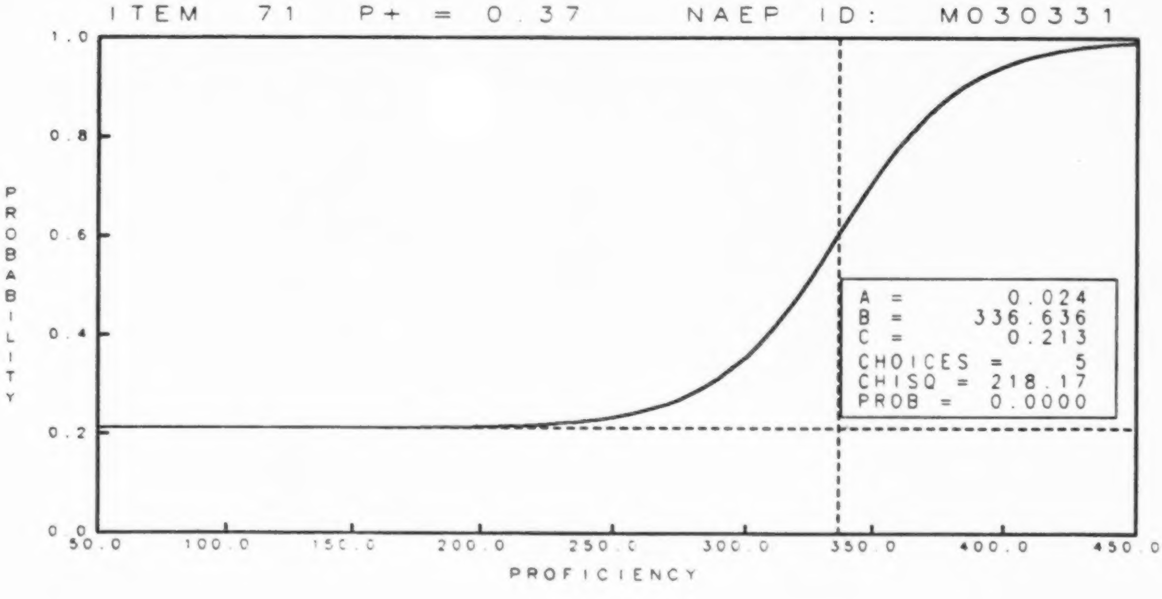
MATH CROSS-SECTIONAL: YEAR 21 - ALL 5 SUBSCALES
BILOG: FINAL ITEM PARAMETER ESTIMATES - TRANSFORMED

SUBSCALE: 1:NUM&OP

BILOG RUN DATE: 10/24/1990 TIME: 16:56:16



ITEM 70 P+ = 0.45 NAEP ID: M029931



ITEM 71 P+ = 0.37 NAEP ID: M030331

TRANSFORMATION PARAMETERS: SLOPE = 50.352 INTERCEPT = 251.719

ICCPLOT VERSION 1.3 DATE: 10/24/1990 TIME: 17:12:46

Appendix M
Listing of Items in Grade-Level Pools in Order of p-Values

GRADE 4
MATHEMATICS ITEMS
PERCENTAGE OF STUDENTS RESPONDING CORRECTLY TO THE 1990 MATHEMATICS ITEMS

Short Text	Total
Find Relative Size Of Numbers	2.3
Complete A Geometric Pattern	8.7
Draw An Obtuse Angle	8.8
Use a Rule To Complete A Chart	14.7
Draw A Geometric Figure	16.7
Solve An Inequality	17.4
Apply Part-Whole Relationship	18.4
Manipulate Numbers	19.7
Read A Scale Diagram	21.2
Divide with A 3-Digit Divisor	22.2
Find Area Of A Rectangle	22.2
Find Perimeter Of A Rectangle	22.6
Read A Ruler	23.5
Estimate Distance on Map	23.9
Visualize A Cube	24.2
Solve Story Problem (Fractions)	24.5
Use A Number Line Graph	25.1
Solve Multi-Step Story Problem	27.4
Visualize Written Statement	28.1
Draw A Geometric Figure	28.5
Solve A Probability Problem	29.0
Draw Geometric Figure	29.6
Apply Concept Of Equality	30.6
Apply Concept Of Area	31.0
Extend A Number Pattern	31.4
Solve Multi-Step Story Problem	32.4
Convert Inches To Feet	32.5
Solve Story Problem (Remainder)	33.3
Find Perimeter Of Rectangle	33.9
Complete A Letter Pattern	34.0
Solve Multi-Step Story Problem	34.9
Apply Properties Of A Cube	35.6
Find Difference In Times	35.6
Use Part-Whole Relationship	36.3
Solve Story Problem (Division)	36.7
Identify Correct Explanation	37.0
Apply Place Value	37.2
Apply Concept Of Perimeter	37.8
Understand When To Estimate	41.3

Short Text	**Total**
Interpret Bar Graph Data	41.6
Identify an Even Number	41.9
Solve Story Problem (Division)	42.5
Interpret Pie Chart Data	43.3
Draw Axis Of Symmetry	43.5
Compare Weights	43.9
Recognize Correct Operation	45.0
Solve Multi-Step Story Problem	45.3
Use A Ruler	45.7
Interpret Reading On A Gauge	46.0
Apply Concept Of Fraction	46.4
Solve Multi-Step Story Problem	47.1
Solve Multi-Step Story Problem	48.6
Identify Solution Procedure	49.3
Identify Parallel Lines	49.3
Represent Words with Symbols	50.1
Apply Place Value	50.2
Complete A Bar Graph	50.7
Determine Greatest Metric Unit	50.9
Add And Divide Whole Numbers	51.0
Solve A Number Sentence	52.1
Identify A Number Relationship	52.4
Solve Story Problem (Multiplication)	52.4
Use A Ruler	55.7
Apply Concept Of Probability	56.0
Solve Ratio Problem	56.0
Solve Story Problem (Multiplication)	56.2
Find Sum Using Number Line	56.4
Apply Properties Of A Square	56.6
Add Whole Numbers	60.0
Find Greatest Distance Between Points	60.1
Determine Missing Fact	60.6
Apply Place Value	61.4
Interpret Decimal Representation	61.4
Apply Transitive Property	61.7
Subtract Whole Numbers	61.7
Solve Story Problem (Multiplication)	61.8
Solve Story Problem (Money)	62.0
Visualize a Geometric Figure	62.0
Read A Graph	63.3
Estimate By Inspection	64.1
Identify Example Of Cylinder	64.7
Solve Story Problem (Reasoning)	65.9

Short Text	Total
Represent Place Value	67.3
Solve Number Sentence (Addition)	69.1
Apply Transformational Geometry	69.4
Solve Number Sentence	70.6
Analyze Volume Relationships	73.0
Interpret Representation Of Fraction	74.2
Multiply Decimals	74.4
Read A Weight Scale	76.2
Extend Geometric Pattern	76.3
Subtract Whole Numbers	76.5
Divide Whole Numbers	76.9
Compare Weights	78.1
Apply Concept Of Probability	78.3
Read Data On Bar Graph	79.7
Write Number Sentence (Multiplication)	79.9
Estimate Distance Given Time	80.3
Determine Largest Number	80.8
Find Greatest Monetary Value	81.3
Subtract Whole Numbers	82.0
Use Order Of Operations	82.1
Multiply Whole Numbers	82.2
Read A Bar Graph	86.1
Add Whole Numbers	88.3
Solve Story Problem (Addition)	88.8
Locate Object On A Grid	89.9
Apply Concept Of Symmetry	91.9
Solve Number Sentence (Addition)	94.0

GRADE 8
PERCENTAGE OF STUDENTS RESPONDING CORRECTLY TO THE 1990 MATHEMATICS ITEMS

Short Text	Total
List Sample Space	10.9
Find An Average	12.3
Solve Story Problem (Conversion)	14.7
Explain Geometric Pattern	14.8
Write Algebraic Expression	14.8
Find A Probability	17.4
Use Least Common Multiple	17.6
Find Percent Increase	17.9
Extrapolate Number Pattern	18.6
Find Width Of A Rectangle	19.0
Find A Median	19.9
Find Total Surface Area	20.3
Interpret Measurement Tolerance	21.4
Identify Perpendicular Segments	21.5
Draw A Line of Symmetry	23.3
Use Scientific Notation	23.8
Apply Pythagorean Theorem	25.3
Order Fractions	27.1
Convert Temperatures	27.8
Apply Pythagorean Theorem	29.2
Fit Equation To Data	29.9
Use Concept Of Midpoint	29.9
Use A Protractor	30.7
Find Divisors Of An Integer	33.6
Find Expected Value	34.0
Recognize Geometric Pattern	34.0
Graph An Inequality	35.2
Read A Scale Diagram	35.4
Apply Concepts Of Exponents	35.7
Locate Point On Graph	36.2
Interpret A Given Rule	36.3
Identify Perpendicular Lines	37.1
Identify Triangle Type	37.4
Add Monomials	38.0

Short Text	Total
Apply Concept of Probability	38.7
Find Ratio Of Side To Perim (Triangle)	40.9
Solve Two-Step Story Problem	41.5
Apply Properties Of A Parallelogram	42.1
Use Similar Triangles	42.5
Find Angle In Triangle	42.6
Relate Equation To Figure	43.2
Apply Concept Of Volume	43.6
Solve Story Problem (Decimals)	43.7
Identify Algebraic Identity	44.0
Interpret Circle Graph	44.1
Identify Coordinates On A Grid	44.4
Solve An Inequality	45.5
Solve A Proportion	45.5
Explain Sampling Bias	46.0
Use Tangrams	46.0
Solve Multi-Step Story Problem	46.2
Use A Rule To Complete A Chart	46.6
Apply Concept Of Average	47.9
Estimate Decimal/Fraction	47.9
Solve Story Problem (Multiplication)	49.2
Solve A Proportion	49.4
Complete A Letter Pattern	49.5
Identify A Number Pattern	49.7
Solve Story Problem (Fractions)	49.7
Convert Fraction To Decimal	50.3
Convert Within Metric System	50.9
Use Tangrams	52.2
Apply Division	53.0
Apply Decimal Place Value	53.8
Visualize A Cube	54.4
Apply Place Value	55.0
Compare Weights	55.0
Solve An Inequality	55.0
Use Percent Greater Than 100	55.1
Draw A Geometric Figure	55.8
Draw Geometric Figure	57.1
Find Probability (Visual Stimulus)	58.1
Use A Number Line Graph	58.6
Apply Ratio And Proportion	58.7

Short Text	**Total**
Apply Properties Of A Cube	58.8
Converts Units Of Time	59.3
Find Perimeter Of Figure	59.4
Apply Transformational Geometry	59.7
Find Checkbook Balance	60.3
Read a Ruler	60.7
Find An Average	61.4
Apply Properties of Geometric Solids	61.9
Interpret A Line Graph	62.1
Apply Part-Whole Relationship	62.8
Find Area Of A Rectangle	63.9
Apply Concept Of Perimeter	64.6
Extend A Number Pattern	65.7
Apply Concept Of Equality	66.5
Solve Story Problem (Remainder)	66.6
Add Two Integers	67.6
Identify A Parallelogram	67.7
Apply Triangle Inequality	68.0
Draw an Obtuse Angle	68.2
Identify 3-Dimensional Shape	69.4
Use A Ruler	69.4
Solve Multi-Step Story Problem	69.5
Complete A Number Sentence	70.5
Apply Place Value	71.0
Interpret Pie Chart Data	71.7
Convert Chart To Circle Graph	72.7
Interpret Bar Graph Data	74.1
Estimate Distance on Map	75.1
Identify A Diameter	75.2
Solve A Probability Problem	75.2
Understand When To Estimate	75.9
Evaluate An Expression	76.6
Relate Equation To Problem	76.9
Solve Multi-Step Story Problem	76.9
Solve a Number Sentence	76.9
Use A Ruler	76.9
Identify Solution Procedure	78.4
Visualize A Geometric Figure	78.4
Convert Decimal To Percent	78.5
Represent Words With Symbols	79.1
Add Whole Numbers	79.7
Apply Transformational Geometry	80.3
Solve Story Problem (Division)	81.7
Solve Story Problem (Multiplication)	81.7

Short Text	Total
Find a Common Factor	82.5
Read A Ruler	82.6
Apply Concept Of Probability	83.0
Identify Measurement Instrument	83.5
Solve Story Problem (Money)	83.5
Subtract Whole Numbers	83.6
Apply Multiplication	84.7
Complete A Bar Graph	85.6
Compare Weights	86.7
Interpret Representation Of Fraction	88.8
Solve An Equation	89.0
Read Data On Bar Graph	89.1
Identify Unit Of Length	90.5
Solve Story Problem (Reasoning)	90.7
Read A Measure On A Scale	91.8
Add Whole Numbers	92.1
Use Order Of Operations	94.1
Use Order Of Operations	94.4
Complete A Geometric Pattern	94.8

GRADE 12
PERCENTAGE OF STUDENTS RESPONDING CORRECTLY TO THE 1990 MATHEMATICS ITEMS

Short Text	Total
Calculate Probability	2.3
Find Volume Of A Cube	3.5
Write Algebraic Expression	8.6
Solve A Quadratic Equation	9.0
Count Combinations	10.3
Write Algebraic Equation	10.7
Find Sine Of Angle	14.9
Apply Interest (Money)	14.9
Sketch A Triangle	15.2
Apply Recent Increase	19.7
Find A Point On A Sine Curve	19.9
Apply Pythagorean Theorem	20.8
Use Trigonometric Ratios	20.8
Explain Application Of Percent	21.8
Find A Median	22.1
List Sample Space	22.1
Apply Area Of A Triangle	24.7
Solve System Of Equations	24.9
Find Coordinate Of Point On Unit Circle	25.0
Solve A Rate Problem	25.1
Interpret Statement	25.4
Find Term Of A Sequence	25.7
Apply Composition Of Functions	25.8
Graph Absolute Value	25.9
Compare Areas	26.8
Explain Geometric Pattern	27.3
Write Algebraic Expression	27.5
Estimate Exponential Growth	27.6
Visualize Intersection In Space	27.6
Use Least Common Multiple	28.5
Find An Average	28.7
Sum Lengths Of Arcs	29.0
Find Total Surface Area	29.2
Apply Scientific Notation	29.7
Draw A Line Of Symmetry	29.9
Find A Probability	30.6
Estimate Circumference	31.3
Solve Quadratic Inequality	33.7
Find Terms In A Sequence	34.3

Short Text	**Total**
Extrapolate Number Pattern	35.7
Describe Graph Of Inequality	35.8
Interpret Measurement Tolerance	36.8
Solve Multi-Step Story Problem	37.7
Find Slope Of A Line	38.9
Solve Area Problem	39.0
Interpret Function Graph	40.6
Explain Application Of Percent	42.1
Apply Pythagorean Theorem	43.2
Substitute And Solve Formula	43.9
Relate Independent/Dependent Variables	44.3
Find Area Of A Square	45.1
Use Scientific Notation In Division	45.3
Convert Liquid Measure	46.3
Apply Pythagorean Theorem	46.9
Approximate Square Roots	47.1
Use Concept Of Midpoint	47.2
Read A Scale Diagram	47.3
Find Side Of Square	47.7
Interpret Function Graph	48.3
Find Percent	49.0
Identify Perpendicular Segments	49.1
Recognize Geometric Pattern	49.3
Find Expected Value	49.7
Interpret A Given Rule	50.2
Interpret Logic Statement	50.9
Supply A Counterexample	51.8
Evaluate A Function	52.2
Compute With Date In Table	52.4
Estimate Height	52.4
Apply Property Of Obtuse Triangle	53.0
Apply Concept Of Volume	53.3
Write A Composite Function	54.6
Identify Triangle Type	56.8
Fit Equation To Data	57.3
Solve Two-Step Story Problem	57.6
Find Range Of Scores	58.4
Convert Decimal To Fraction	59.1
Relate Equation To Figure	59.5
Complete A Letter Pattern	60.4
Apply Concept Of Probability	61.1
Divide Decimals	62.0
Use Signed Number Concept	62.4
Apply Properties Of A Parallelogram	62.8

Short Text	Total
Graph An Inequality	63.0
Solve A Proportion	63.2
Apply Concept Of Percent	63.3
Solve Story Problem (Fractions)	64.6
Solve An Inequality	65.1
Interpret Data In Table	65.1
Find Volume Of A Cylinder	65.3
Recognize Properties Of A Rectangle	65.7
Identify Coordinates On A Grid	67.0
Interpret Pictograph	67.2
Apply Concept Of Average	68.5
Apply Property Of Obtuse Triangle	68.6
Relate Metric To English Units	68.8
Apply Properties Of A Cube	69.9
Explain Sampling Bias	69.9
Find Angle In Triangle	70.2
Use Similar Triangles	70.4
Find Probability (Visual Stimulus)	70.9
Apply Concept Of Perimeter	71.4
Convert Units Of Time	73.8
Evaluate An Expression	74.1
Interpret A Line Graph	74.6
Apply Transformational Geometry	74.8
Use Concept Of Percent	74.9
Identify A Sphere	75.4
Interpret Circle Graph	75.4
Multiply Fractions	75.5
Apply Decimal Place Value	76.1
Apply Properties Of Geometric Solids	76.2
Compare Products (Money)	76.2
Multiply Fractions	76.9
Use A Number Line Graph	77.7
Solve An Inequality	78.6
Compute With Data In Table	79.0
Find Radius (Centimeters)	79.5
Add Monomials	79.5
Apply Concept Of Equality	79.7
Interpret Data In Table	80.0
Read A Ruler	82.7
Solve Multi-Step Story Problem	82.8
Find Checkbook Balance	84.0
Interpret Pie Chart Data	84.4
Apply Transformational Geometry	86.2

Short Text	Total
Find Dividend	86.7
Apply Transitive Property	87.9
Complete A Bar Graph	88.0
Estimate Distance On Map	88.1
Apply Additive Inverse	88.6
Find Verticle Angle Measure	89.0
Interpret Representation Of Fraction	89.2
Identify Solution Procedure	89.4
Solve Multi-Step Story Problem	89.5
Read A Protractor	89.6
Compare Weights	89.8
Solve Story Problem (Division)	89.8
Apply Multiplication	91.0
Interpret Data In Table	91.1
Solve Story Problem (Money)	91.8
Change Percent To Decimal	92.8
Add Whole Numbers	94.1
Read A Measure On A Scale	96.0
Use Order Of Operations	96.1

Appendix N

Acknowledgments

Appendix N

Acknowledgments

It is almost two years since the National Assessment Governing Board first conceptualized the process for setting achievement levels. During that time, literally hundreds of individuals have worked long and hard to implement this landmark initiative of the Board. This report is one of the fruits of those efforts.

The Board would like to thank the National Center for Education Statistics (NCES) staff, particularly Emerson Elliott, NCES Acting Commissioner, Steve Gorman, Gary Phillips, and Eugene Owen for their cooperation and support. This report and much of the replication and validation effort have been funded through NCES, and we are grateful for their continued support of the Board's work.

Educational Testing Service (ETS) played a significant role. Special thanks goes to Archie Lapointe, NAEP Project Director, and Ina Mullis, Deputy Director. We are also grateful to John Barrone, Albert Beaton, Eugene Johnson, and John Mazzeo for their technical wisdom and advice, and assistance with various analyses.

The Board extends its thanks to the many professionals under whose able direction this project proceeded. We owe a debt of gratitude to Ronald K. Hambleton, principal consultant to the project. To Edward Haertel and Robert Forsyth who ably provided technical advice throughout the project, we offer our thanks. Daniel Stufflebeam, Richard Jaeger, and Michael Scriven served on the external evaluation team. George Bohrnstedt of the Trial State Assessment evaluation team, and Robert Linn of the Technical Review Panel served this project well through their ongoing evaluation and advice. We are also grateful to Russell Jones and Mohammad Dirir for their computer analyses related to the standard-setting process.

We wish to thank all those who participated in the standard setting process, and in the replication and validation efforts. Without their dedication and assistance we would have been unable to complete this very important work. To the teachers, supervisors, administrators, and assessment personnel at the state and local levels, to the leaders in business and industry, to parents, school board members, and all others who so ably assisted us, we extend our thanks.

Finally, we wish to thank Paulette Henson, Juanita Taylor, Mary Beth Mason, and Reg Louraine for their word processing and technical contributions to this report, as well as Lisa Hammer and Munira Mwalimu, for their superb handling of all the logistical arrangements required to successfully complete the standard-setting process. We appreciate the work of Sandra Thomas who took responsibility for orchestrating the validation meetings. We are grateful to all the reviewers and editors who provided invaluable suggestions for improving this report.

National Assessment Governing Board
1100 L Street, NW
Suite 7322
Washington, DC 20005–4013

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

FOURTH CLASS BOOK RATE

"...the Board shall ... [identify] appropriate achievement goals for each age and grade in each subject area to be tested under the National Assessment;"

Public Law 100-297